Population Growth, Internal Migration and Urbanisation in Malaysia: Recent and Future trends, 1980 – 2040

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor of Philosophy by:

Muhammad Hakim Bin Danial

July 2019

Acknowledgements

First and foremost, I would like to give my utmost gratitude to God who eases my PhD journey and my life to this very day. I would also like to express my sincere gratitude to Dr Paul Williamson, my main supervisor who continuously supports me to the very end of my PhD for his understanding, patient and vast wisdom. To my co-supervisors, Professor Hill Kulu and Dr Dani Arribas-Bel thank you for being part of my research and all the guidance that was given. To my family, especially my mom (Jamaliah) and my wife (Farah) who relentlessly supports me to this very day despite all the ups and downs. To my friends, Syamsul, Nisa and Nazrith for treating me as a family while in Liverpool. To the School of Environmental Science Postgraduate (SOES PGR) administrator and my colleague (Chris, Leo, Emily, Sarah, and Anna), thank you for being part of my journey. Finally, I am indebted to my sponsors (Ministry of Education Malaysia and Universiti Teknologi MARA) for giving me the opportunity to study at the University of Liverpool.

Abstract

Urbanisation in developing countries has rapidly emerged since 1950 with a somewhat similar experience as developed countries. However, urbanisation has occurred much faster in developing countries. However, in the early 1970s, urbanisation shifted towards counterurbanisation in most developed countries. In contrast, there is little evidence that developing countries will experience counterurbanisation due to their complex nature in terms of historical, economic, and social conditions. To examine the transition process, Geyer and Kontuly (1993) introduced differential urbanisation theory to explain the concentration and deconcentration processes within urban systems (from urbanisation to polarisation reversal to counterurbanisation). Due to rapid urbanisation in Malaysia during the last few decades, this theory fits the aim of this research to examine recent and future population growth, internal migration, and urbanisation in Malaysia from 1980 to 2040. The findings of this thesis show that Malaysia experienced the second urbanisation stage (Intermediate Primate City) since 1980 but shifted towards the final stage (Advanced Primate City) by 2000 due to the shift from rural-urban to urban-urban migration in the capital metropolitan suburban areas. If recent demographic trends persist in the future, the country is projected to remain in the Advanced Primate City stage until 2040. However, there is some indication that this country may experience a polarisation reversal in the future due to the shrinking concentration in capital metropolitan suburban areas and a gradual increase of the concentration in regional metropolitan areas. This thesis makes four original contributions: 1) a rare application of differential urbanisation theory in a developing country context and for the first time in Malaysia; 2) the creation of a new settlement type in Malaysia that can be applied consistently from 1980 onwards and is compatible with differential urbanisation theory; 3) a detailed analysis of socio-economic drivers of internal migration in Malaysia through the application of smaller geographical units and far more socio-economic factors and types of flows; and 4) the determination of a new settlement type and district-level projections of Malaysia's future population using assumptions equivalent to those included in the official national- and state-level projections.

Table of contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Introduction

Urbanisation in developing countries has rapidly increased since 1950 and shares some similarities with the experience of urbanisation in developed countries. On the other hand, there are differences, particularly that urbanisation has occurred much faster in developing countries. According to Jedwab, Christianesen, and Gindelsky (2015), it took more than 100 years, from the eighteenth to the nineteenth century, for developed countries (particularly in Europe) to reach 40 percent urbanisation. In comparison, developing countries (in Asia and Africa) reached the same stage almost twice as fast, from 1950 until 2010. However, in the early 1970s, most developed countries experienced a change in urbanisation patterns—the concentration of the population in metropolitan areas reached its peak, followed by a deconcentration of the population, with small and medium-sized cities seeing higher net migration flows than the largest city (see Argent & Rolley, 2012; Berry, 1980; Champion, 2003; Coombes, Longa, & Raybould, 1989; Fielding, 1982; Halliday & Coombes, 1995; Kontuly & Vogelsang, 1988). Since then, urbanisation has shifted towards counterurbanisation. Counterurbanisation can be interpreted as the movement of the population from a concentrated region to fewer concentrated areas, including movement beyond metropolitan boundaries (Champion, 2003; Coombes et al., 1989; Halliday & Coombes, 1995). Counterurbanisation has arisen for many reasons: clustering job opportunities, access to higher-level services, more choices for housing, the establishment of new towns, stringent planning controls in urban areas, and regional policies (new investment outside of major cities) (Hosszú, 2009).

Urbanisation in developing countries has been rapid. However, instead of industrialisation, the main cause of urbanisation is pressure and constraints in rural areas, which force migration from rural communities to urban areas (e.g., increasing rural poverty and unemployment levels, a surplus of agricultural labour, and cultural factors that force members of rural communities to migrate to urban areas) (Mihermutu, 2011). Furthermore, rural-urban migration in developing countries has also been caused by the broadening of gaps and uneven socio-economic levels, and the spatial segmentation of development between rural and urban areas (Shamshad, 2012). This has led authorities in developing countries to implement various policies and restrictions to prevent massive inflows of migrants to cities (Lall, Selod, & Shalizi, 2006). As of yet, there is little evidence regarding whether developing countries will experience counterurbanisation.

To examine the transition from urbanisation to counterurbanisation, Geyer and Kontuly (1993) introduced differential urbanisation theory to explain the concentration and deconcentration of the population in a temporal sequence within the urban system. The transition between urbanisation stages can be identified when the net migration to one settlement type exceeds that of net migration to another settlement type (Champion, 2005). For example, urbanisation happens when net migration to the largest city exceeds net migration to other cities, while counterurbanisation happens when net migration to small cities exceeds net migration to large and medium-sized cities. Between urbanisation and counterurbanisation, there is a polarisation reversal stage in which medium-sized cities have the largest net migration rather than large and small cities. Differential urbanisation theory combines existing theories (urbanisation, polarisation reversal, and counterurbanisation) into one over-arching theory that has been applied in many developed countries (see Bonifazi & Heins, 2003; Champion, 2003; N. Geyer & Geyer, 2017; Heikkila, 2003; Nefedova & Treivish, 2003; Tammaru, 2003). However, only a few studies have applied the theory in a developing country setting (see Gedik, 2003; Gwebu, 2006; Mookherjee, 2003) and none in the Malaysian context.

During the last few decades, Malaysia has experienced rapid urbanisation. Existing studies show large cities in Malaysia have lost their primacy due to a significant deconcentration of the population since 1970 (see Abdullah, 2003; Hasan & Nair, 2014; Osman, Abdullah, & Nawawi, 2017). The over-arching aim of this thesis is to examine recent and future population growth, internal migration, and urbanisation in Malaysia to consider the extent to which they support or challenge the application of differential urbanisation theory in the Malaysian context. Based on recent evidence, the theory assumption could be useful in explaining recent and possible future urbanisation in Malaysia.

1.2 Background

Historically, the modern urban system in Malaysia was first initiated by the British colonial regime to strengthen their control over and further exploit the country. Urbanisation and population growth in Malaya (the name of the country before Malaysia was formed in 1963) were mainly driven by significant international immigration from China and India. These immigrants were allocated to towns by the British colonial regime due to increasing demand for war-related workers during the Second World War (Lestari, 1997). After the war ended, the growth was entirely sustained through 1960 by natural increases in the population as a result of high fertility levels due to the improvement of nutrition, preventive health programs, and greater accessibility to curative medicine (Hirschman, 1980). However, from the 1960s onward, fertility levels began to decline continuously. This situation was caused by several factors: social change, improved education, women's empowerment in the working sector, and the postponement of marriage and childbearing. Mortality, on the other hand, has been declining since at least the 1950s and continues to do so (Hirschman, 1980).

Besides a natural increase, urbanisation and population growth in Malaysia have also been influenced by rural-urban migration. Resettlement programs imposed by the colonial regime forced rural communities to migrate into new settlements with the aim of denying or blocking insurgent forces from getting support from these communities (Yaakob, Masron, & Masami, 2010). Migration provided opportunities for communities to become involved in commercial, trading, and mining activities in the new settlements. Furthermore, the opening of tin mines encouraged large-scale in-migration of workers, which led to the establishment of more seaports for trading activities. In time, mining and industrial growth were unable to meet the increasing demand for labour caused by rapid rural-urban migration (Yaakob et al., 2010). Not only that, large concentrations of the population in cities had a major impact on urban development and growth and put pressure on the Malaysian government to provide more expenditure for housing, educational, health, and institutional facilities.

Geographically, the distribution of urban centres in Malaysia is uneven. Most are located in high-density areas in the west coast region of peninsular Malaysia and have existed and grown continuously since the colonial period. Major urban centres such as *Kuala Lumpur, Georgetown*, and *Johor Bahru* became the main destination for migrants seeking better economic and social opportunities. However, the primacy of these cities has eroded since the 1980s due to a decline of urban population growth, and the concentration has shifted towards the suburban areas (Abdullah, 2003). The proportion of the population in the largest city, *Kuala Lumpur,* was almost on par with the surrounding suburban areas in the year 2000. The same situation occurred in *Georgetown* and *Johor Bahru*, where areas adjacent to the cities had a larger proportion of the population in 2000. The *Kuala Lumpur Structure Plan 2020 Report* (2003) states that the decline in population growth is one of the main problems in *Kuala Lumpur* (Dewan Bandaraya Kuala Lumpur, 2003). Furthermore, a previous mayor of *Kuala Lumpur* expressed the view that the city is practically dead once office workers leave and return to their homes in the suburbs (Shuid, 2004).

Based on the background review, it is clear that urbanisation in Malaysia has generally been caused by both natural urban population growth and rural-urban migration. However, there are a few problems and questions with this basic outline, which are explained in the next section.

1.3 Problem statement

By reviewing the existing literature on urbanisation in Malaysia, three key problems can be identified: 1) different urban definitions and spatial units; 2) a lack of small scale (sub-state) studies; and 3) no prior studies on possible future urbanisation patterns.

1. Different urban definitions and spatial units

Accounts of urban growth in Malaysia are highly dependent upon the definition and measures of urban areas that are used. The following government agencies use different official definitions: Federal Department of Town and Country Planning in Peninsular Malaysia and Malaysia Department of Statistics. According to the Second National Physical Plan (produced by the Department of Town and Country Planning in Peninsular Malaysia), there are 249 cities in Malaysia in 2010. On the contrary, the Department of Statistics Malaysia outlines only 149 cities. This variance is mainly due to the different definitions and measures used to define urban areas, which results in the production of different information (e.g., urban-rural spatial boundaries, total population, and land area covered).

In 2010, the Malaysia Department of Statistics defined urban areas as:

> *Gazetted areas with their adjoining built-up areas, which had a combined population of 10,000 or more at the time of the Census 2010 or the special development area that can be identified, which at least had a population of 10,000 with at least 60 % of the population (aged 15 years and above) were involved in non-agricultural activities.*
>
> *Department of Statistics of Malaysia (2010)*

The Federal Department of Town and Country Planning in Peninsular Malaysia, on the other hand, use a similar definition but includes extra criteria for defining urban areas: 1) population density of 50-60 people per hectare; and 2) the existence of urban infrastructure and facilities. The department outlines seven levels of the urban hierarchy in the *National Urbanization Policy*: national growth conurbation (population of more than 2.5 million), regional growth conurbation (1.5 to 2.5 million population), sub-regional growth conurbation (0.5 to 1.5 million population), state growth conurbation (0.3 to 0.5 million population), district growth conurbation (0.1 to 0.3 million population), major settlement centre (30,000 to 0.3 million population), and minor settlement centre (10,000 to 30,000 population). The national growth conurbation consists of the national capital and the largest city in Malaysia (Kuala Lumpur), the national administration centre (Putrajaya), and surrounding suburban areas.

2. Lack of small-area studies

There are few urbanisation studies focused on Malaysia that apply small-area units in their analysis. Small-area studies commonly involve administrative units below the national and state levels, such as local authority areas, townships, municipalities, postal areas, and residential districts. In the Malaysian context, the administrative unit that follows the state is the district, which is followed by *mukim* and the local authority area (LAA). Table 1.1 shows the spatial units used for the analysis of existing studies:

Table 1. 1: Urbanisation studies in Malaysia

Author(s)	**Spatial unit(s)**
Hasan and Nair (2014), Mohd Jali (2009)	National, State, and certain District
Lestari (1997)	State and certain LAA
Aiken & Leigh (1975), Hirschman (1976), Masron, Yaakob, Mohd Ayob, and Mokhtar (2012), Sendut (1966)	LAA
Abdullah (2003, 2012), Abdullah and Mohd (2009), Osman et al., (2017), Osman, Nawawi, and Abdullah (2009)	Certain District and Certain LAA

Based on Table 1.1, no studies have focused on all district and *mukim* levels. The main reason the LAA unit is commonly used in urbanisation studies is that they consist of urban and rural boundaries, which is ideal for examining urbanisation. However, only basic information is available about LAA populations (e.g., total population, population by gender, age groups and ethnicity) and migration (e.g., total migration from urban and rural areas). This provides insufficient data to examine urbanisation in a comprehensive way (e.g., types of urban areas, migrants' origins and destinations, births and deaths in urban and rural areas). In contrast, this information is widely available for other small-area units (district and *mukim*).

3. There are no studies on future urbanisation trends in Malaysia

So far, no studies have focused on the future of urbanisation in Malaysia. Projections for the future is an important tool that provides guidance to policymakers and helps them plan interventions to guide societal and economic development. However, existing projections focus only on population and migration changes for large-area units (state and national levels) from 2010 through 2040. Recent studies have shown that urban sprawl has been a major problem in Malaysia since the 1980s due to rapid suburbanisation in metropolitan regions caused by uncontrolled urban development and growth (Abdullah, 2012; Hasan & Nair, 2014). The increase in large-scale urban development projects, the concentration of migrants in metropolitan areas, and problems with urban sprawl may have major impacts on urban development and growth. These issues put pressure on the Malaysian government to control development and provide more expenditure for housing, infrastructure, and amenities. Therefore, projections that focus on future possibilities for urbanisation could help the Malaysian government implement more effective policies to mitigate both current and future issues associated with urbanisation.

The aim and objectives of this thesis focus on addressing the three problems identified above.

1.4 Aim and objectives

The aim of this thesis is to investigate recent and future population growth, internal migration, and urbanisation in Malaysia from 1980 to 2040 through a small-area analysis. Five objectives were formulated to achieve this aim:

1. To review existing theoretical perspectives on population growth, internal migration, and urbanisation.
2. To develop a new urban-rural classification for Malaysia based on urbanisation theory.
3. To investigate patterns of population growth, internal migration, and urbanisation in Malaysia over the recent period (1980-2010).
4. To identify and explain the determinants of internal migration in Malaysia from 1980-2010.
5. To project future population growth, internal migration, and urbanisation over the period from 2015 to 2040.

1.5 Thesis structure

The structure of this thesis corresponds to the five thesis objectives, as illustrated in Table 1.2

Table 1. 2: Objectives and corresponding chapter(s)

No.	Objective	Corresponding chapter(s)
1	To review existing theoretical perspectives on population growth, internal migration and urbanisation.	Chapter 2: Literature review
2	To develop a new urban-rural classification of Malaysia based on urbanisation theory.	Chapter 3: Developing an urban-rural classification
3	To investigate patterns of population growth, internal migration and urbanisation in Malaysia over the recent period (1980-2010).	Chapter 4: Population growth and urbanisation in Malaysia, 1980-2010 Chapter 5: Internal migration and urbanisation in Malaysia, 1980-2010
4	To identify and explain the determinants of Malaysia internal migration from 1980-2010.	Chapter 6: Determinants of internal migration in Malaysia 1980-2010

5	To project future population growth, internal migration and urbanisation over the period 2015 to 2040.	Chapter 7: Future population growth, internal migration and urbanisation in Malaysia, 2015-2040

Chapter 2 tackle the first objective by reviewing the existing literature on topics related to the research. This is an important preliminary chapter because it provides a foundation for examining urbanisation, population growth, and internal migration by reviewing existing theories and empirical evidence. The subsequent chapters apply these theories to examine the topics of population growth, internal migration, and urbanisation in the Malaysian context.

As mentioned in Section 1.3, since the definitions of urban areas differ between government agencies, and because there is limited data for existing urban-rural units, Chapter 3 tackles the second objective by developing a new urban-rural classification for Malaysia based on the theoretical perspectives introduced in Chapter 2. The chapter first introduces Malaysia's geographical context (existing spatial units and settlement hierarchies) before outlining the process of developing a new urban-rural classification. The resulting classification represents an advance on existing classifications because it 1) is directly linked with differential urbanisation theory, 2) uses geographical building blocks that permit linkage to the specific aspects of available census data, and 3) covers the entire period of 1980 to the present.

Chapter 4 and Chapter 5 use the new urban-rural classification to address the third objective through observing recent trends in population growth, internal migration, and urbanisation in Malaysia from 1980-2010. These chapters make a major contribution by testing differential urbanisation theory and its applicability to the developing world since most prior applications of the theory are based on Western-based experience.

Chapter 6 investigates the determinants of internal migration in Malaysia. This chapter uses spatial interaction models to explore the nature of aggregate migration flows between the places of origin and destinations. In the process, a series of models are developed to identify the socio-economic factors that influence these flows, including the urban-rural nature of the origin and destination. The chapter concludes by developing a final model that best explains the main causes of internal migration in Malaysia from 1980-2010.

Chapter 7 addresses the final objective of this thesis by projecting future urbanisation trends. This is achieved by projecting future births, deaths, and internal migration using a district-level cohort-component model. This is in contrast with currently available official projections, which operate at the national and state levels only and thus have nothing to say about Malaysia's future urbanisation trajectory. The chapter concludes by evaluating the success of the differential urbanisation theory in explaining the future urbanisation process in Malaysia.

Finally, Chapter 8 summarises and discusses all of the key findings in the thesis, identifying the key contributions made and drawing an overall conclusion concerning the links between the differential urbanisation theory and Malaysia's recent and possible future urbanisation pathway. Recommendations are also made regarding future work required to expand on the work embodied in the thesis.

1.6 Conclusion

Since 1950, urbanisation has emerged much faster in developing countries than what has been observed historically in developed countries. Further, in the early 1970s urbanisation shifted towards counterurbanisation in most developed countries. This raises questions regarding whether developing countries will follow the same path. To date, there is little evidence on this matter. There is a possibility that developing countries may not experience counterurbanisation due to their complex nature in terms of historical, social, economic, and cultural conditions. To understand the urban transition process, Geyer and Kontuly (1993) introduced the differential urbanisation theory to explain the urban concentration and deconcentration process within the urban system in a temporal sequence. Recent studies have shown that major cities in Malaysia have lost their primacy due to the large deconcentration of the population since 1970. However, few urbanisation studies in the Malaysian context have been undertaken at the sub-state level. For these reasons, this thesis sets out to investigate recent and future population growth, internal migration, and urbanisation in Malaysia from 1980 to 2040 at the small-area (sub-state) level.

By applying differential urbanisation theory to Malaysia, this thesis makes the following original contributions:

i. Rare application of differential urbanisation theory in a developing country context and for the first time in Malaysia.
ii. Creation of a new settlement type for Malaysia geography that can be applied consistently from 1980 onwards and that is compatible with differential urbanisation theory.
iii. A detailed analysis of socio-economic drivers of internal migration in Malaysia by considering smaller geographical units and numerous socio-economic factors and types of migration flows.
iv. The first settlement-type and district-level projections of Malaysia's future population using assumptions equivalent to those included in the official national and state-level projections.

In the following, Chapter 2 reviews the academic context of this thesis.

Chapter 2

Literature review

2.1 Introduction

To understand population change, internal migration, and urbanisation in Malaysia, it is necessary to first review what is already known about how these processes operate elsewhere. Therefore, this chapter reviews what is currently known about the nature, causes, and consequences of population change, migration, and urbanisation, in both the developed and developing world, to identify which aspects may be applicable to the Malaysian experience. The conclusion reached is that differential urbanisation theory provides a useful lens through which to view all of these processes. The chapter is organised as follows. First, the meaning of the term 'urbanisation' is considered. This is followed by a discussion of the links between demographic change and urbanisation. In more industrially advanced societies, it has been argued that the urbanisation process is complete and that such societies have experiences 'polarisation reversal' (dispersal from larger to medium-sized cities) and even 'counterurbanisation (dispersal to small-sized cities and rural areas). These ideas are explored in Sections 2.4 and 2.5. Following this, Section 2.6 introduces differential urbanisation theory, which attempts to draw together the processes of urbanisation, polarisation reversal, and counterurbanisation into a single over-arching theory. Differential urbanisation theory is not without its critics, and Section 2.7 reviews some of the competing theories related to the urbanisation process. Although these criticisms may offer some helpful additional insights, it is concluded that the underlying basis of differential urbanisation theory remains sound. Sections 2.8 and 2.9 thus review the lessons that can be learnt from previous attempts to apply differential urbanisation theory in a range of developed and developing countries.

Finally, the chapter concludes by summarising the key criticisms of differential urbanisation theory and its relative utility.

2.2 Urbanisation

First of all, this section provides a brief overview of definitions of urbanisation patterns, including how they differ between developed and developing countries. Generally, urbanisation is defined as the increase of the proportion of the urban population resulting from rural-urban migration, the natural increase of urban population, expansion of urban boundaries, annexation of surrounding areas to urban centres, or the creation of new urban centres (Hasan & Nair, 2014; Hosszú, 2009; Tacoli & Satterthwaite, 2015). However, urbanisation cannot be distinguished based on the expansion of urban land if the expansion rate is higher than urban population growth because in this case it is associated with declining settlement density, also known as urban sprawl (Abdullah et al., 2009; Mcgranahan & Satterthwaite, 2014). The definition of an urban area, on the other hand, varies between countries. Despite these differences, common criteria of urban areas include the size of the population, population density, and administrative status (Tacoli & Satterthwaite, 2015). Finally, the terms urban population growth and urbanisation are often used interchangeably in studies and are commonly misinterpreted (Tacoli & Satterthwaite, 2015). For example, urbanisation does not occur if the urban and rural population grow simultaneously.

From a historical point of view, urbanisation in Western countries accelerated during industrialisation (eighteenth to nineteenth centuries) but decelerated after the early 1970s due to counterurbanisation in response to growing individual affluence (Fielding, 1982; Jedwab et al., 2014). Urbanisation in developing countries, on the other hand, emerged rapidly after 1950. In some ways, this urbanisation pattern resembles the experience of developed countries, but there are also differences. Urbanisation has occurred much faster in the developing world. According to Jedwab, Christianesen, and Gindelsky (2015), it took more than a century (from the eighteenth to the nineteenth century) for developed countries (particularly in Europe) to reach 40 percent urbanisation. In contrast, developing countries (in Asia and Africa) reached the same stage almost twice as fast, between 1950 and 2010. Furthermore, rapid industrialisation was not the main cause of urbanisation. Rather, urbanisation was caused by technological and institutional change, or by pressure and constraints in rural areas (e.g., agricultural crises, famine, droughts, and poverty) that forced residents of rural communities to leave and migrate to urban areas (Fox, 2012; Mihermutu, 2011).

To expand the discussion of this topic, and because the focus of this thesis is on examining the relationship between demographic change and urbanisation, the next section discusses urbanisation from a demographic perspective.

2.3 Urbanisation from a demographic perspective

Following the previous section, this section reviews urbanisation from a demographic perspective, focusing on natural population increase (Section 2.3.1), rural-urban migration (Section 2.3.2), and urban residential mobility (Section 2.3.3).

2.3.1 Natural population increase

Natural population increase is generally identified as the difference between fertility and mortality levels and has been argued to be one of the main factors of urbanisation (Dyson, 2011; Lestari, 1997). This can be interpreted in two ways: 1) if the birth rate is higher than the death rate, the population will increase; and 2) if the death rate is higher than the birth rate, the population will decrease.

One of the famous theories used to examine fertility and mortality trends is demographic transition theory. Demographic transition theory was first introduced by American demographer Frank Notestein, who explained fertility and mortality change as a set of sequences (Kirk, 1996). It is known that the first demographic transition in developed countries began roughly in the 1780s during the industrialisation revolution and ended in 1960 (Khan, 2008). Reher (2004) reviewed the theory from a global perspective by observing mortality and fertility trends by characterising several countries into different groups (*Forerunners, Followers, Trailers,* and *Latecomers*). The results of his work confirm the validity of the theory, as all groups experienced a similar pattern: a mortality decline came before a fertility decline. According to Kurek et al. (2015), the second demographic transition occurred in 1990 as a result of a negative population change, negative net migration, and negative natural increase in urban core areas. The transition is characterised by four features: 1) fertility level below replacement levels; 2) increasing mean age of mothers when giving birth; 3) declining number of marriages; and 4) increasing divorce rate.

Similarly, Wilson (2011), in his studies on demographic convergence, examined mortality by investigating life expectancy in four regions (Soviet Union, developed countries, developing countries, and Africa) from 1950-2010. The growing and linear trends of life expectancy between developed and developing regions reflect growing access to health facilities and medicine. Following the increase of life expectancy and decrease of mortality, fertility trends declined in developed and developing regions. Before the 1970s, fertility trends in these regions showed a stable pattern, with a huge gap between the fertility rates. However, there was a significant decrease of fertility in developing regions after 1970, shrinking the gap for most economically advanced countries in East Asia and Latin America. Furthermore, the rising and stable trends of fertility in developed regions since 2000 will further speed up the convergence process. The main cause for this is exposure to and use of modern contraceptives, postponement of childbearing and marriage, increasing abortion rates, and higher levels of female education and employment (Ernestina, 2002; Hirschman, 1980).

While most studies on natural population increase have been focused on the national level, a few have examined natural population increase in local contexts. Recent studies in developed countries have shown that fertility levels relate to the size of settlement: the bigger the settlement, the lower the fertility (Kulu, 2013). In other words, rural areas have higher fertility rates than urban areas. The main reason for this is the tendency to have more children in smaller settlements than in larger settlements and housing type and location preferences (see Kulu, 2005; Kulu, Boyle,

& Andersson, 2009; Kulu & Vikat, 2007). In terms of mortality differences between urban and rural areas, existing studies indicate that different countries display different patterns of mortality. For example, rural areas in the United States tended to have higher mortality than urban areas for 1999-2014 (Moy et al., 2017). On the contrary, urban areas in European countries during the eighteen and nineteenth centuries had higher mortality than areas in the surrounding countryside (Woods, 2003). More detailed study in the UK found that there are differences between urban and rural areas for most causes of death, mortality rates are similar for some causes (e.g., cancer, circulatory disease) (Gartner, Farewell, Dunstan, & Gordon, 2008).

Note here that demographic transition theory has little to say about migration because it is mainly focused on national-level population processes instead of local-level processes. However, once the focus is on the local level, migration flows become very important in helping to shape local population change.

2.3.2 Rural-urban migration

Rural-urban migration plays an important role in the urbanisation process. It has been argued that cities face more challenges than rural areas (Awumbila, 2017). In 2014, more than 50 percent of the global population lived in urban areas, and it is predicted that this figure will continue to increase to 66 percent in 2050 (United Nations, 2014). Cities are known to be drivers of urbanisation, resulting from the concentration of business, government, and national economic activity as well the provision of better infrastructure and facilities than those in rural areas (United Nations, 2014). However, broadening gaps in terms of socio-economic conditions and the spatial segmentation of development between urban and rural areas has resulted in uneven social and economic growth and has led to rapid rural-urban migration (Shamshad, 2012). The decision to migrate from a rural to an urban area highly depends on individual income and living cost differential arises from the spatial segmentation of labour and the capital market (Saracoglu & Roe, 2004). This is true in many places; for example, recent studies in China indicated that the increase of long-distance migrants in cities is influenced by income disparities between high and low economic growth regions (Chan, 2012). Generally, rural-urban migration is caused by push and pull factors. Push factors are characterised by high unemployment rates, low wages, few job opportunities, a lack of infrastructure and basic amenities, and poverty while urban pull factors are associated with perceived high employment opportunities, decent wages, and better facilities.

To date, there are numerous studies on rural-urban migration in the literature that have focused on a variety of factors in different locations. For example, rapid urbanisation in India has been found to be caused by the heavy influx of migrants from deprived areas, which leads to socio-economic and environmental problems in urban areas (Shamshad, 2012). In Botswana, it was found that education level influences migration decisions because of the expectation of higher earnings in urban centres (Gwebu, 2006). The experience in Botswana may relate to the hypothesis in Harris and Todaro's two-sector model, which assumes that migrants' decisions are based on rational economic reasons (Harris & Todaro, 1970). In this model, rural migrants are assumed to be attracted by the expectation of higher earnings and are willing to accept lower wages and the risk of unemployment in urban areas. Rapid rural-urban migration in Ethiopia, on the other hand, was seen during the post-revolutionary period in the late 1970s due to high employment in urban areas, rural poverty, economic transitions into a capitalist system, political instability, and government resettlement policies (Mihermutu, 2011). The situation in Ethiopia is closely related to dependency theory, which suggests urbanisation is a result of the transition of the economic structure towards a global capitalist system (Peng, Chen, & Cheng, 2011). Finally, in terms of cultural factors, men in Kenya are considered to be breadwinners, thus encouraging them to migrate to support their families (Mihermutu, 2011).

The rapid growth of cities in developing and less developed countries, caused by huge influxes of migrants, results in various problems in urban areas (e.g., housing shortages, slums, homelessness, increasing crime and poverty) (Yaakob et al., 2010). Furthermore, according to Rodriguez and Rowe (2018), rural-urban migration can also cause a rise in the number of working-age adults, a decline in local sex ratios, and downgraded educational levels in large cities. This situation has led the government in many countries to impose alternatives that seek to diminish huge rural-urban migration flows. Many developing and less developed countries evidently attempt to discourage rural-urban migration through the implementation of policies and restrictions to prevent massive inflows of migrants to cities (e.g., adapting apartheid or a nativist system, resettlement programmes, industrial decentralisation, equal investment across settlements, and rural development schemes) (Lall, Selod, & Shalizi, 2006).

While rural-urban migration studies typically involve migration behaviour in a regional or district context, urban residential mobility, on the other hand, is more focused on a local or residential context. This is discussed in the next section.

2.3.3 Urban residential mobility

According to Bell et al. (2015), residential mobility can be characterised by semi-permanent changes of residential address and involves short-distance movement or intra-urban migration. On the contrary, internal migration usually involves long-distance movement beyond the 'daily time-space geography' and permanent change of address (White, 2016). One of the earliest studies was done by Long (1988), who compared stages of internal migration in multiple countries. He suggested two fundamental approaches for examining internal migration: comparison of residential mobility and distance. There were, however, few residential mobility studies at that time.

Nevertheless, the topic has grown in importance in recent years. Several studies have found that urban residential mobility is closely related to natural population increase in urban areas and is associated with fertility levels. Generally, residential mobility encourages people to expand their families by having more children and allowing them to attain a better living environment, mainly in terms of housing. Fertility levels are known to be higher in suburban areas than in the urban core and are based on the size of the urban area: the larger an urban area, the lower the fertility level (Kulu & Boyle, 2009; Kulu et al., 2009; Kulu & Washbrook, 2014; Kurek et al., 2015). Furthermore, location and housing type play important roles in changes to fertility: residential relocation and living in a larger house or a 'family-friendly environment' evidently raise fertility levels (Kulu, 2005; Kulu & Vikat, 2007).

Moreover, higher fertility levels are more prevalent among migrants than non-migrants but have little impact due to their small share of the population (Kulu & Boyle, 2009; Kurek et al., 2015).

Residential mobility patterns, however, are different in developing and less developed countries. In the context of Latin American countries, the decline of net-inflows in the largest metropolitan cities are caused by urban problems (e.g., high crime, housing shortages) due to the poor governance and limited financial resources (Rodríguez & Rowe, 2018). Furthermore, due to housing scarcity and economic constraints, this situation may also lead to the mushrooming of informal housing (e.g., slums, squatters) in urban areas. Rapid natural population increase and large streams of in-migration to urban areas can lead to housing shortages and thus growth of uncontrolled, informal settlements (Kasarda & Crenshaw, 1991). Commonly, residential mobility in developing countries is evidenced from the movement of migrants to different types of shelter in urban areas, such as relatives' houses, rental housing, or 'live-in servitude', which, in the end, equate to living in informal housing (Kasarda & Crenshaw, 1991).

In summary, both natural population increase and migration significantly affect the urbanisation process of a country. However, at some point, as urban areas start to mature and age, the urbanisation process begins to slow down and spatial deconcentration begins. The population in medium-sized cities (located close to but not contiguous with a metropolitan region) will eventually start to grow (Geyer & Kontuly, 1993). This transition process is known as polarisation reversal.

2.4 Polarisation reversal

Before discussing on polarisation reversal, it is helpful to understand the urbanisation transition process. According to Geyer and Kontuly (1993), during the advanced stage of urbanisation, rapid population growth and agglomeration economies in the core metropolitan region transform the spatial structure to become inefficient and costly. Rising land values and congestion costs encourage the decentralisation of economic activities to satellite centres within the core region. Expanding job opportunities may divert new migrants into these cities rather than to the primate city. However, this intraregional decentralisation pattern does not resemble polarisation reversal because the primate city tends to grow faster than other cities. In time, the dispersion of economic activities into peripheral regions becomes more efficient associated due to agglomeration economies and scale economies, which are reflected in increasing population and income, expanding markets, infrastructure development, and lower input costs. Furthermore, the dispersion process may also be fuelled by continued rapid expansion and negative externalities in the core metropolitan region (increasing living costs, increasing congestion, and housing cost pressure), thus accelerating the de-industrialisation process and inducing the decentralisation of migrants, indicating the beginning of polarisation reversal. Polarisation reversal is known as the turning point in the spatial polarisation pattern and growth of the national economy and leads to spatial decentralisation and urban deconcentration (Richardson, 1980). So far, most developed countries have already experienced the polarisation reversal stage. There are also clear indications that polarisation reversal may occur in developing and less developed countries (Geyer Jr & Geyer, 2016; Gwebu, 2006; Mookherjee, 2003).

2.5 Counterurbanisation

In time, as medium-sized cities can no longer accommodate population growth or growing urban development, the population will disperse towards the countryside or rural areas (Geyer & Kontuly, 1993). This phenomenon is known as counterurbanisation. Counterurbanisation is interpreted as the process of demographic deconcentration beyond a metropolis and its suburban areas (Champion, 2003; Coombes et al., 1989; Fielding, 1982; Halliday & Coombes, 1995). Counterurbanisation arises for many reasons: moving out of large cities to smaller towns, clustering of job opportunities, access to higher-level services, the availability of housing, demographic factor (e.g., life course events such as when to have more children), economic growth, establishment of new towns, stringent planning controls in urban areas, regional policies (new investment away from major cities), restructuring of the manufacturing industry, and the expansion of motorways (Hosszú, 2009). Counterurbanisation, however, is not merely the relocation of urban residents to the countryside; it also includes cultural and social factor. For example, counterurbanisation may happen when a migrant population feels the lifestyle or culture in rural areas is better or equivalent to urban lifestyles (Mitchell, 2004). Historically, growth rates and the concentration of the population in non-metropolitan or countryside areas reached its peak in the early 1970s in most developed countries. Counterurbanisation was evident in the United States, Britain, France, Australia, Italy, and West Germany at this time from major population shifts towards non-metropolitan regions (See Argent & Rolley, 2008; Berry, 1980; Champion, 2003; Coombes et al., 1989; Fielding, 1982; Kontuly & Vogelsang, 1988). Most non-metropolitan regions experience the same migration pattern: age-selective out-migration followed by the in-migration of a large number of elderly to the countryside.

2.6 Differential Urbanisation Theory

Differential urbanisation theory draws together urbanisation, polarisation reversal, and counterurbanisation into one over-arching theory. Geyer and Kontuly (1993) introduced the theory to explain the concentration and deconcentration process of a population in a temporal sequence. The transition between urbanisation stages can be identified by a series of 'clean breaks'. A clean break is a situation in which net migration to one settlement type exceeds net migration to another settlement type, resulting in a change of the urbanisation stage (Champion, 2005). For example, polarisation reversal occurs when net migration to medium and small cities exceeds net migration to the largest city while counterurbanisation occurs when net migration to small cities exceeds that to large and medium-sized cities. Figure 2.1 and Table 2.1 explain differential urbanisation theory model in detail.

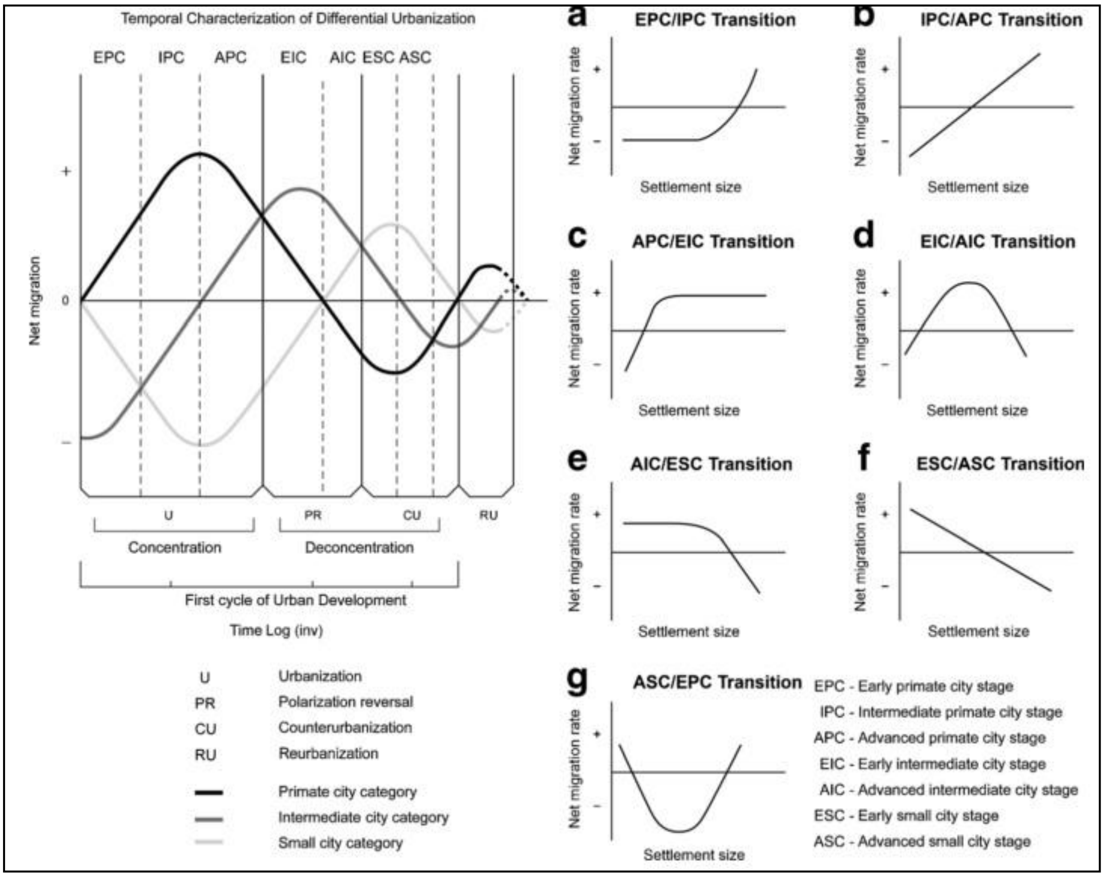


Figure 2. 1: Differential urbanisation theory model

Table 2. 1: The sub-stages of differential urbanisation theory

Cycle	Stage	Sub-stages	Explanation
First cycle	Urbanisation	Early primate city	The primate city has the highest net rural-urban migration.
		Intermediate primate city	The population in primate city is rapidly growing and early suburbanisation occurs due to net migration.
		Advanced primate city	In time, agglomeration diseconomies and decentralisation towards suburban centres form a metropolitan region. Population in the primate city maintains its dominance compared to cities of other sizes.
	Polarisation reversal	Early intermediate city	As the primate city starts to mature and age, the growth rate begins to slow down and spatial deconcentration starts. The population in intermediate-sized cities (located close to but not contiguous with a metropolitan region) starts to grow. The primate city is still gaining population in absolute terms but is starting to lose in terms relative to intermediate-sized cities. Further, the population in suburban centres is increasing faster than in the primate city.

		Advanced intermediate city	The population in intermediate-sized cities continues to increase rapidly (at a smaller scale) and population in the metropolitan region declines in absolute terms (the primate city loses more population than suburban areas).
	Counterurbanisation	Early small city	Similar to the previous stage (early intermediate city), the growth in intermediate-sized cities starts to slow down, and the growth in the primate city continues to decline. The population is now deconcentrating towards small cities.
		Advanced small city	The population in small cities grows faster than in both primate and intermediate-sized cities.
Second cycle	Re-urbanisation		In time, due to limitations, small cities can no longer expand or accommodate the population and urban growth. As the urban system matures, migration rates slightly decline and intermediate-sized cities generally do not supersede large cities nor do small cities supersede intermediate-sized and large cities. Thus, the relationship between cities becomes fixed, and they are restricted to certain roles based on hierarchal order. Furthermore, the hierarchical relationship between cities is based on function rather than their absolute size in the urban system. The second sequence of urbanisation begins when the primate city has the highest growth from natural population increase and in-migration rather than rural-urban migration.

Source: Summary based on the theoretical concept of differential urbanisation theory by Geyer and Kontuly (1993).

The three main urbanisation stages in Figure 2.1 are distinguished based on the relative growth of cities of different sizes: large, intermediate, and small. Within these stages, there are several sub-stages (Table 2.1). Urbanisation is evident when the population in primate city grows faster than in cities of other sizes due to rapid rural-urban migration. Polarisation reversal begins when net in-migration to intermediate-sized cities exceeds that to the primate city. Finally, the urban system reaches maturity by entering the counterurbanisation stage, during which small cities start attracting significant numbers of migrants compared to both intermediate-sized and primate city. In time, small cities can no longer expand or accommodate the population and urban growth. As the urban system matures, migration rates slightly decline, and intermediate-sized cities generally do not supersede large cities nor do small cities supersede intermediate-sized and large cities. Thus, the relationship between cities becomes fixed, and they are restricted to certain roles based on a hierarchal order. During this time, the second cycle sets in when net migration to the largest cities once again exceeds that to small cities. The r-urbanisation phase sets in, and large cities once again become more significant and concentrated at the expense of intermediate-sized and small cities.

The theory also assumes that heterogeneity of the population results due to contradictory migration patterns and motivations of different subpopulation groups in response to different social and economic pressures (Geyer Jr., Geyer, Plessis, & Eeden, 2012). A mainstream migration pattern occurs when large subpopulations dominate net migration patterns while sub-stream migration displays a contrary pattern. Mainstream and sub-stream migration patterns occur independently and can be seen in different stages of migration at any given point in time. In developed countries, mainstream migration generally involves moving towards primate cities during the urbanisation phase. As the urban system matures, the deconcentration process starts to set in with the outward migration of the population from primate cities. This model assumes that there is a strong linkage between the movement of capital and labour. De-industrialisation in intermediate-sized cities, followed by a population deconcentration, is the first indication that the country is entering the polarisation reversal phase.

As mentioned earlier, the heterogeneity of the population leads to different migration motivations. *Productionism* and *environmentalism* are regarded as important economic and social motivations in both concentration and deconcentration movement in differential urbanisation theory (Geyer Jr. & Geyer, 2015). *Productionism* is based on economic motivations from social agglomeration economies (e.g., due to more employment opportunities, better wages, lower transportation costs, more profitable markets, and cheaper labour). *Environmentalism* involves motivations driven by the pull factors of non-metropolitan regions towards increased quality of living (e.g., less congestion, lower living costs, better education and public services, and fewer safety concerns) (Geyer Jr et al., 2012).

Generally, migration rates marginally decline during cycles of concentration and deconcentration. As the urban system matures, migration rates decline slightly, and intermediate-sized cities generally do not supersede large cities, nor do small size cities supersede intermediate-sized and large cities. Thus, the relationship between cities becomes rigid, as cities take on particular functional roles based on the hierarchal order (Geyer and Geyer Jr., 2015, p.4). Urban systems tend towards long-term stability in the hierarchical relationship between cities. The stability of the urban system of a country, however, can be achieved during the first-phase cycle only, and other phases may not be experienced, as postulated by the theory.

However, differential urbanisation theory is not the only possible way to examine urbanisation, and there is a significant amount of literature on the topic. Therefore, the next section reviews some competing views, theories, and evidence on the urbanisation process.

2.7 Alternative theories

Scholars continuously attempt to understand and interpret the complex nature of the urbanisation process and patterns across regions. However, theories that were formed based on Western experiences are insufficient to explain urbanisation in most developing or less developed countries due to the countries' different and complex natures. The combination of historical, social, economic, and cultural conditions have created dissimilar patterns of urbanisation to those suggested by Western-based theories and experience. Although the theories are not universally applicable, they still provide a great foundation for explaining the topic. Therefore, the following paragraphs review some competing views, theories, and evidence on urbanisation.

2.7.1 Urbanisation theories

Self-generated or endogenous urbanisation theory suggests two different conditions for viewing the urbanisation process: 1) the population engaged in non-agricultural activities is sustained by the production of surplus products and 2) social achievement of large communities allows them to be 'socially viable and stable' (Peng et al., 2011). The theory emphasis that urbanisation is caused by the transition of the population from rural to urban areas, driven by rapid industrialisation processes. As highlighted in much of the urbanisation literature, this was historically proven in the majority of Western countries during the Industrial Revolution, when rapid industrialisation accelerated cities' growth (Khan, 2008). Classically, the transition of the population from rural to urban areas resulted in the growth in urban labour markets (Gottmann, 2004). In contrast, labour demand in the agricultural industry in rural areas declines, leading to increased unemployment, which forces members of rural communities to leave (Hosszú, 2009). The assumption of the endogenous urbanisation theory is inadequate to describe urbanisation as it only considers the rural-urban transition of the population due to industrialisation.

One example is from a study done by Fox (2012), who argued urbanisation in sub-Saharan Africa occurred differently than what is postulated by existing urbanisation theories. He argued that urbanisation in South Africa is actually driven by technological and institutional change instead of industrialisation during the preindustrial era. He claimed the shortage of food and limited access to disease prevention methods decelerated population growth. The technological and institutional changes introduced by colonizers, however, reduced these constraints and limitations. Urbanisation in African countries is also driven by agricultural crises, famine, droughts, and poverty, which force people to move to urban areas (Hosszú, 2009; Mihermutu, 2011). After the post-war period, urbanisation was influenced by huge demographic changes (i.e., a major decline in mortality rates and increases in fertility and life expectancy rates) instead of economic growth.

The situation in South Africa relates to the modernisation theory, which explains urbanisation in a broader context than just highlighting rural-urban movement and industrialisation factors. The theory assumes economic and social development in developing countries can be attained from diffusion of the growth of institutions, trade, and global companies based in developed countries (Lotfi, 1998). At some point, development in less developed areas emerges, driven by an increase of interaction and integration with developed areas. The linear pattern involves transformation from a traditional into a modern and more rational society, which is associated with a population shift from rural to urban areas. Modernisation starts in the largest cities and is later dispersed to settlements according to the hierarchy sequence (Lotfi, 1998).

Modernisation theory outlines three concepts: 1) the initial state of urbanisation and development is determined by early of modernisation, 2) technology development is the main factor of urbanisation, and 3) patterns of urbanisation between developed and developing countries are united through cultural dispersal despite uneven urban development (Kasarda & Crenshaw, 1991; Peng et al., 2011). Many past empirical studies have proven the first concept: urbanisation was driven by the existence of technology introduced by colonisers of African countries (Fox, 2012). Besides the influence of colonisation, surplus from agricultural activities of traditional societies led to the creation of various social and economic patterns and government bureaucracy, as well as urban systems. For the second concept, social change is more prevalent due to the application of technology rather than society's organisation itself. Finally, the third concept explains that, despite the complex nature of and diversity between countries, technology is logically known to replace traditional social organisation with its own ‘institutional matrix’ (Kasarda & Crenshaw, 1991). However, this theory fails to consider other complex factors in the less developed world: economic and spatial disparities, ‘elite power’, differences in policies, global effects, and urban conflicts (Lotfi, 1998).

Following the failure of the modernisation theory, the dependency theory suggests urbanisation is a result of the transition and growth of the economic structure towards a global capitalist system. The theory assumes that a capitalist development pattern exists in the form of social organisation and is characterised by social inequality and uneven development and that urbanisation is internally driven by technology and population dynamics but constrained by external factors (Peng et al., 2011). The presence of global capitalism in domestic urbanisation, particularly in developing countries, has resulted in an imbalance in growth between urban and rural areas, the rise of urban conflicts, rapid rural-urban migration, and increasing centralization of activities and urban primacy in major cities (Lotfi, 1998). According to Lotfi (1998), in the Latin American context, dependent urbanisation occurred during the post-colonial period when countries integrated the international division of labour by supplying ‘primary commodities for the consumption and production’ of core cities. This theory explains development in developed and less developed countries is mutually important, and current economic and social systems are caused by colonial and capitalist growth.

2.7.2 Migration theories

Neoclassical theories of migration, on the other hand, are closely related to urbanisation theories and were classically adapted from Ravenstein’s migration law, Zipf’s gravity model, the Harris-Todaro two-sector model, and Zelinsky’s mobility transition theory. Ravenstein (1885) hypothesized seven migration laws, which Greenwood (1997) summarised: 1) most migrants move only short distances, commonly to cities; 2) cities that grow rapidly tend to be populated by migrants from proximate rural areas; 3) out-migration is inversely related to in-migration; 4) a major migration wave will generate a compensating counter-wave; 5) those migrating long distances tend to move to large cities; (6) rural people are more likely to migrate than urban people; and 7) women are more likely to migrate than men.

Zipf (1946) later introduced a gravity model that was built upon several Ravenstein laws and assumed a higher volume of migration and larger populations of places of origin and destination communities (Bodvarsson & Van den Berg, 2013). The assumption was made based on wage opportunities of the population in the place of origin that exceeded the current wage earned, causing people to migrate to other locations (Borjas, 1994). In time, as the population in the place of origin rises, the number of people choosing to migrate also rises. Hence, the population in a destination increases and there will be more employment opportunities, which will thus attract more migration.

The Harris-Todaro two-sector model assumes the potential migration decision is based on rational economic reasons (Harris & Todaro, 1970). Rural migrants are assumed to be attracted by the expectation of higher earnings and willing to accept lower wages and the risk of unemployment in urban areas. Further, migrants still prefer to migrate if they think there is a possibility of earning more income in the future (Mihermutu, 2011). The two-sector model also focused on the decision-making process from an individual perspective. The model assumes that migrants are likely to get a job and access to information and contacts based on the length of their stay (Jali, 2009). However, the heterogeneity and complexities in urban sectors at different levels may cause problems with the application of this model. According to Young (2004), the labour market may be stratified by gender, ethnicity, or race, which may distort the employment structure. Not only the urban sector but also the rural sector has the same problem, where migrants from rural communities come from different backgrounds and cultures. Another problem with this model is its assumption that expected income is the sole factor for people migrating.

While previous migration theories are more focused on migration, Zelinsky's mobility transition combines all components of change (fertility, mortality, and migration) to explain demographics and the migration process within a set of frameworks in a temporal sequence (Zelinsky, 1971). He assumed the transition process is mainly due to the modernisation of societies, which can be identified in five phases: 1) premodern traditional society; 2) early transitional society; 3) late transitional society; 4) advanced society; and 5) future super-advanced society. During the first phase, he hypothesized that there is little residential migration and limited circular migration accompanied by high fertility and mortality. The onset of urbanisation begins in the second phase with major rural-urban migration, including international and circular migration. However, while fertility levels remain high, mortality rapidly declines. In the third phase, rural-urban migration maintain its dominance but less so than before. This time, both fertility and mortality decline continuously. In the fourth phase, the migration pattern shifts towards urban-urban migration instead of rural-urban migration. Fertility and mortality trends are now stable and are similar. Finally, in the fifth phase, there is no plausible explanation for fertility and mortality, but most migration involves intra-urban and inter-urban residential and circular migration.

Life-course migration theory on the other hand explores the relationship between different transition of life-course and mobility. Young adults tend to have the highest mobility, which then slowly declines with increasing age, sometimes increasing again for those with young children and of retirement age (Bernard, Bell, &

Edwards, 2014). The transition includes people leaving school, beginning higher education, entering the labour force, forming unions, and childbirth (Bernard et al., 2014) and is vary across space. According to Kulu and Washbrook (2014), mobility results in more childbirth in smaller settlement than in larger settlement. One of the reason is due to desire to live in larger house or in a 'family-friendly environment' (Kulu, 2005; Kulu & Vikat, 2007). The mobility encourages people to expand their families by having more children and allowing them to attain a better living environment. Although the transition of life-course and migration display a universal pattern, the ages (particularly the young adults) of which migration occurs differ across countries (Bernard et al., 2014). This was proven in studies done by Bell and Muhidin (2009) that examined cross-national comparison of internal migration of 25 countries. The results of their studies show the migration of young adults in Asian countries are strongly concentrated in the early 20s. On the country, the migration in European and Northern are more concentrated at older ages and is widely spread across age group.

Although the competing theories help increase the understanding of some aspects of the links between population change, migration, and urbanisation, differential urbanisation theory still provides a robust framework for investigating the main issues highlighted in this thesis – as has been mentioned in Chapter 1, the concentration of population in major cities in Malaysia has weakened since the 1980s and shifted towards surrounding cities. This phenomenon fits with differential urbanisation theory assumptions that explain different stages of concentration and deconcentration of population in the urban system. Therefore, the next two sections review previous applications of differential urbanisation theory in case-study countries, first in the developed world and then in the developing world.

2.8 Differential urbanisation in developed countries

As mentioned in Section 2.6, differential urbanisation theory is formed from the combination of several theories, debates, and experiences in the Western world. According to Fielding (1982), urbanisation in Western European countries in 1950 displayed a positive correlation between net migration and settlement size (large settlements had higher net migration compared to smaller settlements). In the 1980s, however, the pattern changed from positive to negative, whereas smaller and medium- sized settlements had higher net migration compared to the largest settlements. The change in migration patterns during 1950-1980 indicates a shift from urbanisation towards counterurbanisation in the majority of countries of Western Europe (Halliday & Coombes, 1995). This section reviews relevant case studies that have adapted differential urbanisation theory for developed countries.

2.8.1 Completion of the urbanisation cycle

To date, Finland is the only country that has completed the first cycle of urbanisation process as postulated in differential urbanisation theory (e.g., from urbanisation to polarisation reversal to counterurbanisation) while most other countries remain in the polarisation reversal stage. According to Heikkila (2003), urbanisation in Finland began before 1940 and lasted until the mid-1950s due to the high net-inflow of migrants in the capital city of Helsinki. One of the main factors of rapid urbanisation in Finland was the industrialisation and reconstruction of cities after the post-World War II period. Most migrants were young people (particularly ages 15 to 24) who migrated in search of jobs or to have a better livelihood. After the 1950s, the urbanisation pattern shifted towards polarisation reversal with the highest net-inflows in medium-sized cities rather than in large cities. In the mid-1970s, the country shifted towards counterurbanisation with negative net-inflow in all cities due to a large out-migration of the population towards countryside areas. Aside from cheaper costs, the urban-rural movement was also motivated by villagers' efforts to attract families who wished to live in the countryside by arranging housing plots, improving the level of services in areas, and launching campaigns to persuade people to return to their home region. Counterurbanisation, however, came to an end after large cities once again received a high concentration of the population during the 1990-1998 periods due to the improvement of employment opportunities in these cities.

2.8.2 Achievement of counterurbanisation

Champion (2003) tested differential urbanisation theory for Great Britain for 1901-1991 by applying the functional urban regions developed by Halliday and Coombes (1995) from the use of local labour market areas (LMMAs). An LMMA is a geographical unit for urban regions that consists of built-up areas, the hinterland, and the local labour market. Champion highlighted that the geographical units used must concern the whole functional units in the urban centre instead of just their restricted boundaries. He found that the main reason for counterurbanisation in Britain is the changing nature and distribution of economic activity. De-industrialisation and economic restructuring led to rapid movement to settlements lower in the hierarchy. This was evident from a rapid decline in manufacturing employment in major cities in Britain for two decades beginning in 1970 and 1980 (Champion, 2003). Other than that, the service-based sector had also grown rapidly in intermediate-sized and smaller cities outside the major cities' conurbation.

Historically, the outward movement in Britain began in the early nineteenth century with the construction of railways and the use of motorised transport (Champion, 2003). Furthermore, the implementation of urban containment policy (the restriction of physical growth, the merging of larger cities, and control of development) through the establishment of new towns away from major cities to accommodate overspill development and population also fuelled the counterurbanisation process (Champion, 2003). Moreover, the rapid decline in household size (small families, marital breakdown, and more young people living independently) and a major shortage of housing required the government to expand the housing stock to keep up with demographic change (Geyer, 2018). Due to limited urban capacity, the government pushed the excess urban population to lower-level settlements. Finally, more affluent or educated people sought out better living conditions while those who were less wealthy remained concentrated in the large cities to survive (Hosszú, 2009).

In terms of ethnicity, almost all ethnicities in Britain experienced a common counterurbanisation pattern instead of race-based movement, contradicting the 'white flag' or white movement portrayed by previous research. Both, whites and minorities have similar proportions of flow to most destinations (Simpson & Finney, 2009). However, recent migration studies have shown that ethnicity does play a significant role in the counterurbanisation process. According to Sapiro (2017), white British display a strong counterurbanisation pattern to countryside areas, followed by Jews and those of Sikh ethnicity.

In West Germany, counterurbanisation began in 1977 and reached maturity in 1980 from the negative correlation between German's net migration and regional population (Kontuly & Vogelsang, 1988; Kontuly, Wiard, & Vogelsang, 1986). Counterurbanisation during 1982-1984 was associated with the movement of citizens and foreign workers towards small cities. Fluctuating demand for foreign workers during the 1970s and 1980s was caused by a decline in national economic growth and an increase of the unemployment rate. This situation was due to restrictive policies in 1973 controlling the influx of foreign workers, which resulted in large out-migration from large cities. Counterurbanisation in West Germany was more prevalent among people aged below 18 years and above 30 years while other age groups displayed weak urbanisation trends. Most young adults (18-30 years old) moved towards metropolitan areas and smaller urban areas while other age groups moved towards suburban areas outside metropolitan boundaries and rural areas (Kanaroglou & Braun, 1992). Clearly, migration by those aged between 18 and 30 years was the result of production-oriented movement based on the availability of employment and educational opportunities. For migrants aged 30 and over, the reasons for migration appear to be a combination of employment consideration and preference for specific amenities by middle-aged adults and the elderly. Furthermore, migrants under 18 years old display similar migration patterns as their parents age 30 years and over (Kanaroglou & Braun, 1992; Kontuly & Vogelsang, 1988).

Counterurbanisation in Australia began in the 1970s from the large in-migration of former metropolitan areas dwellers to non-metropolitan areas (Argent & Rolley, 2008). The deconcentration of the population from metropolitan areas occurred mostly in highly remote areas surrounding major cities and in coastal regions (Argent & Rolley, 2008; Hugo, 2002). The characteristics of the population differ based on accessibility and mobility; the proportion of the youth, highly educated, and unemployed population increased along with remoteness (Hugo, 2002).

However, large in-migration and the existing ageing population in non-metropolitan areas created major problems such as a lack of health facilities and related services (Argent & Rolley, 2008).

According to studies by Tammaru (2003) and Tammaru, Kulu, and Kask (2004), counterurbanisation in Estonia was evident after the Soviet period in the 1990s from the large emigration of people in large cities. Before the 1990s, large cities were sustained by immigrants who worked in the industrial sector while Estonian citizens were focused on the agricultural sector. However, the transition during the post-Soviet period in the 1990s led to large emigration flows of migrants returning to their home country and hence counterurbanisation. Counterurbanisation in Estonia also resulted from under-urbanisation. Limited urban development in large cities (e.g., housing shortages) discouraged people from moving into these cities. Furthermore, due to food shortages, the government imposed various policies (e.g., increasing wages and prices of agricultural products) to improve agricultural productivity, which led more people to live in rural areas. The shift from an industrial to a service-based economy was also slow because large cities were surrounded by rural dominated by agricultural activities and low-density housing (Tammaru, 2003; Tammaru et al., 2004).

2.8.3 Attainment of polarisation reversal

Urbanisation in Italy during 1921-1971was evident from rapid population growth in large cities (Bonifazi & Heins, 2003). Similar to Finland's experience, the distribution of economic activity and industrialisation in large cities after the post-World War II period in the 1950s led to major depopulation in rural and low-density areas (Coombes et al., 1989). In the mid-1970s, the country shifted towards polarisation reversal due to the decline of rural-urban migration and population growth in major cities, as well as de-industrialisation towards medium and small cities. In the 1990s, the growth of these cities was further influenced by interregional and international migration due to exhaustion of local labour. Overall, Italy has not yet encountered counterurbanisation and is still within the polarisation reversal stage.

In South Korea, rapid urbanisation also began after the post-World War II period from high population growth in the capital city of Seoul during 1955-1960 (Rii & Ahn, 2002). This was mainly because of the return of citizens who evacuated during the war alongside existing urbanites and refugees who remained in the city. The implementation of development plans by the South Korean government after the post-war period through economic restructuring and industrial development fuelled mass rural-urban migration in major metropolitan areas during 1960-1970. Due to rapid development and economic growth, metropolitan areas became saturated, causing many urban problems (e.g., traffic congestion, lack of social welfare, insufficient infrastructure, urban sprawl) (Rii & Ahn, 2002). The Korean government implemented various decentralisation policies and programs (e.g., green belts, decentralisation of industries, development of new towns, and tax deductions) to counteract these problems and encouraged decentralisation (Henderson, 2002). The country reached the polarisation reversal stage from a significant decline in population growth in major metropolitan areas starting in 1970 (Kim & Han, 2014). Furthermore, the decline of the employment share in major metropolitans, along with an increase of the employment share in rural and small cities during the 1980s, is a possible indication that the country was moving towards counterurbanisation (Henderson, 2002). However, there is no clear evidence that the country is experiencing counterurbanisation.

Contrary to other developed countries' experience, people in South Korea prefer to stay in or migrate to urban areas rather than migrating to rural areas upon retirement despite many policies imposed by the government to encourage them to emigrate (Kim & Han, 2014). This is mostly because of high accessibility to health services and living amenities. Similarly, this situation commonly occurs in developing and less developed countries, with governments implementing various policies (e.g., resettlement schemes, industrial decentralisation, and nativist systems) to discourage rural-urban migration (Lall et al., 2006).

2.9 Differential urbanisation in developing countries

To date, differential urbanisation theory has mainly been applied in the context of developed countries. To date, the theory has only been applied to these developing countries: India, South Africa, Botswana, and Turkey (Gedik, 2003; Geyer Jr. & Geyer, 2016; Mookherjee, 2003). Based on empirical evidence, the founder Geyer and Geyer (2017) acknowledged that the theory may or may not be applicable to developing and less developed countries and there will probably be significant differences in urbanisation patterns. For example, de-industrialisation or decentralisation towards rural area is the main factors of population deconcentration in developing countries. This is evident from efforts to discourage and prevent massive inflows of migrants in cities through the implementation of various policies, which include adoption of apartheid or nativist systems, resettlement programmes, industrial decentralisation, equal investment across settlements, and rural development schemes (Lall et al., 2006).

2.9.1 Achievement of counterurbanisation

Gedik (2003) tested differential urbanisation theory for Turkey for 1955-2000 and found that the country experienced a slightly different urbanisation pattern in the early stage but follows the same pattern as developed countries in subsequent years. He argued that the urbanisation process in Turkey did not follow the sequence as postulated in the theory and suggested including a 'pre-concentration' stage before the urbanisation stage. The main reason for this is that small cities in Turkey had higher population growth than large cities during the early stage of urbanisation during 1955-1960. He assumed the high population concentration in small cities was due to intra-provincial rural-urban migration, with the lack of a transportation system and communication facilities the main causes for this. Turkey entered the urbanisation stage during 1960-1975, with large cities having the highest population growth compared to medium and small cities. Turkey then entered the polarisation reversal stage during 1975-1980, with the highest population growth in medium-sized cities. The last decade considered, 1990-2000, was considered to be a transitional period between polarisation reversal and counterurbanisation, with population growth in small cities exceeding growth in large and medium-sized cities.

Geyer Jr. and Geyer (2015, 2016) tested differential urbanisation theory for South Africa for 1996-2011 by disaggregating the subpopulations into several categories. Historically, urbanisation in South Africa began with the migration of minorities (the white population) into urban areas due to the destruction of farms and property during the World War II period while the majority population (the black population) largely remained in rural areas. Migration was later influenced by large-scale mining activity associated with industrial development, which attracted a significant proportion of the population (black and white) to urban areas after the post-war period, despite the implementation of apartheid policies. Earlier studies by Geyer (2003) found that the country may have entered the polarisation reversal phase beginning in 1950 due to rapid population growth in regional centres compared to primate cities and small towns. This growth pattern, however, more closely resembles suburbanisation than polarisation reversal due to rapid population growth in satellite towns. Population deconcentration at that time was an indication of secondary sub-stages of urbanisation and not polarisation reversal.

A significant change in migration trends and counterurbanisation in South Africa can be seen since the end of the apartheid era in 1994. Mainstream movement showed weak deconcentration patterns from declining growth rates of the total population in large cities and increasing growth rates in medium-sized cities. Similarly, the dominant subpopulation (blacks and lower-skilled individuals) also experienced the same migration pattern and growth rates. In contrast, the sub-stream movement of the non-dominant (whites and highly skilled individuals) subpopulation showed strong deconcentration patterns from high growth rates in intermediate-sized cities and negative growth in large and small cities. These situations are consistent with the transition process from urbanisation to the polarisation reversal phase outlined in the theory. Further, even though the patterns differ between subpopulations, the trajectories are similar.

Changing migration and urbanisation patterns also relate to different types of motivations, as postulated in the theory. The mainstream population in South Africa exhibited a production-oriented motivation while the sub-stream population exhibited an environmental-oriented motivation (Geyer Jr. & Geyer, 2015, 2016). The decentralization of the sub-stream population (whites and highly skilled individuals) deflected the centralisation of the mainstream population (blacks and low skilled individuals), which led to spatially decentralised urban settlement patterns in primate cities (Geyer Jr. et al., 2012). This was evident from diseconomies due to saturation and high competition in major urban labour markets, which thus drew the population to migrate into medium and small cities.

2.9.2 Attainment of polarisation reversal

Besides South Africa, a study was conducted for its neighbour Botswana. According to Gwebu (2006), urbanisation in Botswana began in the early of 1960s due to the relocation of its capital accompanied by the expansion of existing and new mining towns. This resulted in significant rural-urban movement as a result of employment opportunities in mining centres and the construction sector in towns. Mainstream movement was clearly the result of production-oriented movement based on differential urbanisation theory. Environment-oriented movement was also evident from the establishment of residential areas away from the primate core along with a good transportation system. In addition to the primate cities, suburban areas and regional centres saw significant growth from the development 'spill-over' effect during the 1971-1991 periods. Growth was also fuelled by closures and slow growth in mining centres during 1981-1991. Within the same period, the implementation of decentralisation policies by the government had significant impacts on population growth in sub-regional centres. This growth may have also been influenced by a severe drought, which forced the population in isolated areas to move into larger settlements. The saturation limit and diseconomies in core areas, however, led to the decentralisation of industries and commercial properties. Population growth in core areas began to decline while growth in other areas increased (especially sub-core areas), indicating the country reached the polarisation reversal stage during 1991-2001.

Mookherjee (2003) tested the theory for India during 1961-1991 by an analysis of the national and subnational levels. At the national level, there was higher population growth in large cities than in intermediate and small cities during the 1961-1981 period. The shift towards polarisation reversal began in 1981-1991 when the population in intermediate-sized cities grew more rapidly than in large cities (Jain, Siedentop, Taubenbock, & Namperumal, 2013; Mookherjee, 2003). Population deconcentration in India, however, may not adequately reflect the cyclic process of differential urbanisation theory. At the subnational level, most states remained at the urban concentration stage for two decades from 1961-1981, and none moved towards the deconcentration stage in the subsequent decades. These states appear to have maintained their position during the 30-year period instead of moving to the next urbanisation stage (e.g., polarisation reversal or counterurbanisation). Mookherjee argued that examination of the 30-year period from 1961-1991 may not be sufficient to justify the whole deconcentration process in India due to limitations of India's census data.

On the other hand, an extended study for the 1971-2001 period by Mookherjee and Geyer (2011) found slightly different results: medium-sized cities saw a rapid increase in population while large cities experienced the opposite during 1991-2001 (note that earlier studies captured polarisation reversal beginning in 1981). The authors, however, did not mention why previous and recent studies have captured different urbanisation patterns. One possible reason is that other works may have used a different scale of spatial units or different classifications and may have adjusted settlement types. Changing urbanisation patterns in India were mainly due to the effectiveness of various programmes and policies during the post-independence period aimed at balancing settlement size and population growth. One such policy limited the concentration in large cities by encouraging concentration in other cities through infrastructure development and the development of transportation networks (Mookherjee, 2003; Mookherjee & Geyer, 2011; Seto, 2011). This situation is evident from the decline of employment growth in the large city of Delhi during 1971-2001 indicating development dispersal towards intermediate-sized and small cities (Jain et al., 2013).

There is no clear evidence that India will experience the next deconcentration process (counterurbanisation) even though the country has gone through polarisation reversal. Unlike cities in Western European countries, the inadequacy of physical infrastructure and the lack of institutional capacity may result in re-urbanisation instead of counterurbanisation and thus this situation will further stress the resources in large cities, which will require policy restructuring to improve infrastructure delivery and capacity building (Jain et al., 2013).

2.10 Conclusion

This chapter reviews what is currently known about the nature, causes, and consequences of population change, migration, and urbanisation in both the developed or developing world to identify which aspects are applicable to the Malaysian experience. The conclusion from this review is that differential urbanisation theory provides a useful lens through which to view all of these processes. Although other related theories help with a better understanding of some aspects of the links between demographic changes and urbanisation, differential urbanisation theory provides a robust framework for investigating these issues. The combination of related theories and the experience of developed countries (with urbanisation, polarisation reversal, and counterurbanisation) led to the formulation of differential urbanisation theory, which provides a general framework of the cyclical processes in urban systems. Empirical research is crucial to assess the validity of proposed theories and to address issues or questions that arise concerning their future application. The theory has been tested and discussed over the past decades in the context of several developed countries, thus confirming its validity. It also has been applied for developing countries, and the results clearly indicate that these countries experienced a similar process as developed countries. The success of the model is demonstrated in Table 2.2.

Table 2. 2: List and comparison of empirical studies

Countries	Method(s) used	Urbanisation stage	Author(s)
Finland	Population growth and net migration by macro and micro level	Completion of the urbanisation cycle	(Heikkila, 2003; Vartiainen, 1989)
The United Kingdom and Great Britain	Population growth and net migration by regions, districts, gender and age groups	Counterurbanisation	(Champion, 2003, 2005; Coombes et al., 1989; Halliday & Coombes, 1995; Rees, Durham, & Kupiszewski, 1996)
West Germany	Correlation by total population, age groups, citizen and foreign net-migration		(Fielding, 1982; Kanaroglou & Braun, 1992; Kontuly & Vogelsang, 1988)
Australia	Population growth by accessibility/ remoteness of metropolitan and non-metropolitan regions		(Argent & Rolley, 2008; Hugo, 2002)
Estonia	Net migration between metropolitan and non-metropolitan regions and by settlement size		(Tammaru, 2003; Tammaru et al., 2004)
Turkey	Population growth by settlement size		(Gedik, 2003)
Italy	Population growth and net migration by settlement size	Polarisation reversal	(Bonifazi & Heins, 2003; Coombes et al., 1989; Fielding, 1982)
India	Population growth by national and regional level		(Jain et al., 2013; Mookherjee, 2003; Mookherjee & Geyer, 2011)
South Africa	Population growth by subpopulation groups		(Geyer, 2003; Geyer Jr & Geyer, 2016; Geyer Jr et al., 2012; Gwebu, 2006)
Botswana	Population growth by settlement size		(Gwebu, 2006)
South Korea	Population growth by settlement size		(Fielding, 1982; Kim & Han, 2014; Rii & Ahn, 2002)

Finland is the only country that has completed the first cycle of urbanisation and is moving towards the first phase of the second cycle (re-urbanisation). However, the majority of other developed countries studied have only reached the counterurbanisation stage: Britain, France, West Germany, Australia, and Estonia. There are unclear signs in these countries, however, that they will follow the urbanisation sequence by moving back to an earlier stage or skipping to another stage. Countries that have gone through the polarisation reversal phase include Italy, India, Turkey, South Africa, Botswana, and South Korea. Turkey, however, experienced a slightly different pattern at the early stage but later reached the polarisation reversal stage. Based on the discussion of the literature and empirical evidence, the use of differential urbanisation theory is preferred as the main theoretical framework to investigate the concentration and deconcentration processes of population and migration in the Malaysian context.

However, there are several shortcomings in the application of this theory. First, a country does not necessarily follow the urbanisation sequence postulated by the theory, although this has been empirically proven in some countries. The founders of differential urbanisation theory, Geyer and Geyer (2017), admitted in a recent article that the theory may or may not be applicable to developing and less developed settings, as there may be significant differences in urbanisation patterns. For example, the decentralisation of development or de-industrialisation towards smaller settlements and rural areas are possible main factors of population deconcentration, instead of common factors such as retirement or improvement of individual economic and social capacity (e.g., through wealth and more education) as in most developed countries. This is evident from efforts by governments of developing countries to discourage and prevent massive inflows of migrants to cities by implementing various policies (e.g., adapting apartheid or a nativist system, resettlement programmes, industrial decentralisation, equal investment across settlements, and rural development schemes) (Lall et al., 2006).

Second, it is important to include rural areas in the model (since the theory focuses only on interactions between cities) because more segregation can be seen between rural and urban areas than between urban areas of different sizes (Tammaru, 2003). Third, some of the empirical studies used population growth as the main indicator to explain the urbanisation process, namely those in India, South Africa, Botswana, South Korea, and Turkey (refer to Table 2.2). Changing population growth could be concerned just as much with the changing settlement radius as with changing levels of rural-urban migration. Fourth, the concepts of *Productionism* (economic motivation) and *Environmentalism* (social motivation) introduced in the theory have not been properly tested. Most studies explain urbanisation factors based only on the historical experience of a country instead of quantifying the results. A possible key reason for this is that there is no specific method and a lack of suitable data to analyse the causes of urbanisation.

The next chapter develops a new urban-rural classification for Malaysia that allows differential urbanisation theory to be tested in the country, while simultaneously addressing some of the shortcomings of the theory.

Chapter 3

Developing a new urban-rural classification

3.1 Introduction

As discussed in the previous chapter, the aim of differential urbanisation theory is to explain the concentration and deconcentration processes of populations in a temporal sequence for different settlement sizes: largest city, intermediate-sized cities, and small cities. However, there are no specific guidelines provided in the theory to differentiate cities except for the requirement that they are located independently from each other (Geyer & Kontuly, 1993). Previous studies have used various methods for defining cities because different countries used different measurements and definitions of urban and rural areas. For example, Champion (2003) tested differential urbanisation theory for Great Britain by applying the functional urban regions developed by Halliday and Coombes (1995) using of local labour market areas (LMMAs). An LMMA is a geographical unit for urban regions that consists of physically built-up areas, the hinterland, and local labour market. Mookherjee (2003), on the other hand, tested the theory for India by analysing urban concentration and deconcentration at the national and state levels. In response, this chapter develops a new urban-rural classification for Malaysia to enable the application of differential urbanisation theory approach to the country. Before development of this new urban-rural classification, Section 3.2 introduces the spatial units currently used to explain the country's geographical background. The Malaysian government established the existing urban-rural classification, but the information provided about it is quite limited. Therefore, Section 3.3 explains the process of fitting the existing classification system with the theoretical approach to build a new urban-rural classification. Finally, Section 3.4 concludes the important findings of this chapter. The outcome of this chapter is a new urban-rural spatial unit that is used in the subsequent chapters.

3.2 Existing spatial unit classification in Malaysia

Malaysia is a Southeast Asian country bordering Singapore, Brunei, Thailand, Indonesia, and the Philippines. The country is divided into two regions: the west coast and east coast regions: The regions are subdivided into states, which are subdivided into districts, which in turn are subdivided into *mukim*. Some of the states are governed by the federal government (*Kuala Lumpur, Putrajaya*, and *Labuan*) while others are governed by state governments. As of 2010, there were 16 states (including the federal states), 144 districts, and 932 *mukim*. Figure 3.1 shows the spatial boundaries between the states, districts, and *mukim*.

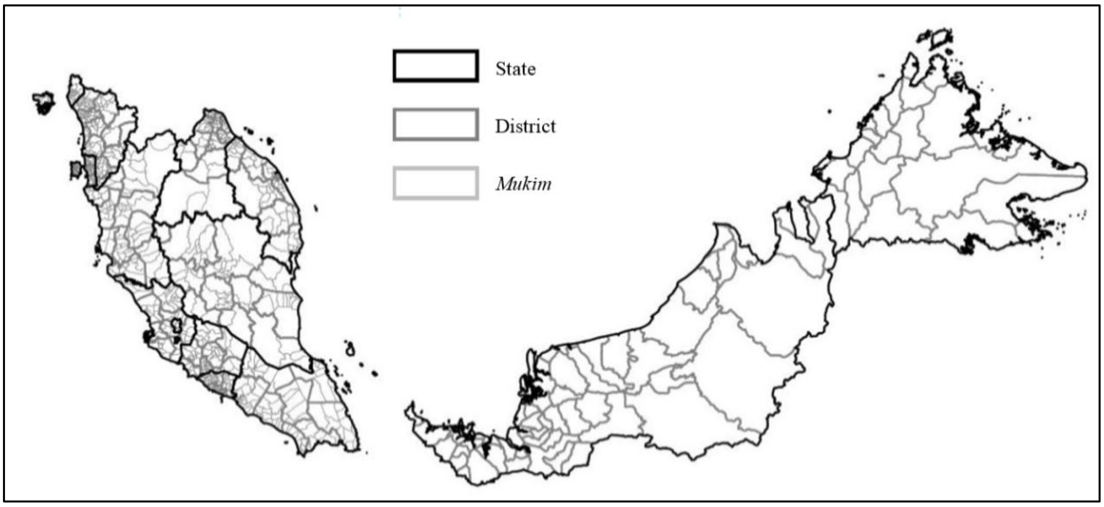


Figure 3. 1: State, district, and *mukim* boundaries in Malaysia, 2010

Another way of spatially subdividing Malaysia is to split it into urban and rural areas. Urban-rural boundaries depend highly on the definitions and measurements used and thus change over time. The existing urban-rural boundaries in Malaysia vary between government agencies. Figure 3.2 and Figure 3.3 shows urban-rural boundaries created by the Federal Department of Town and Country Planning in Peninsular Malaysia and the Malaysia Department of Statistics, respectively.

Figure 3. 2: Hierarchy of cities and their boundaries in peninsular Malaysia outlined by the Federal Department of Town and Country Planning in Peninsular Malaysia, 2006

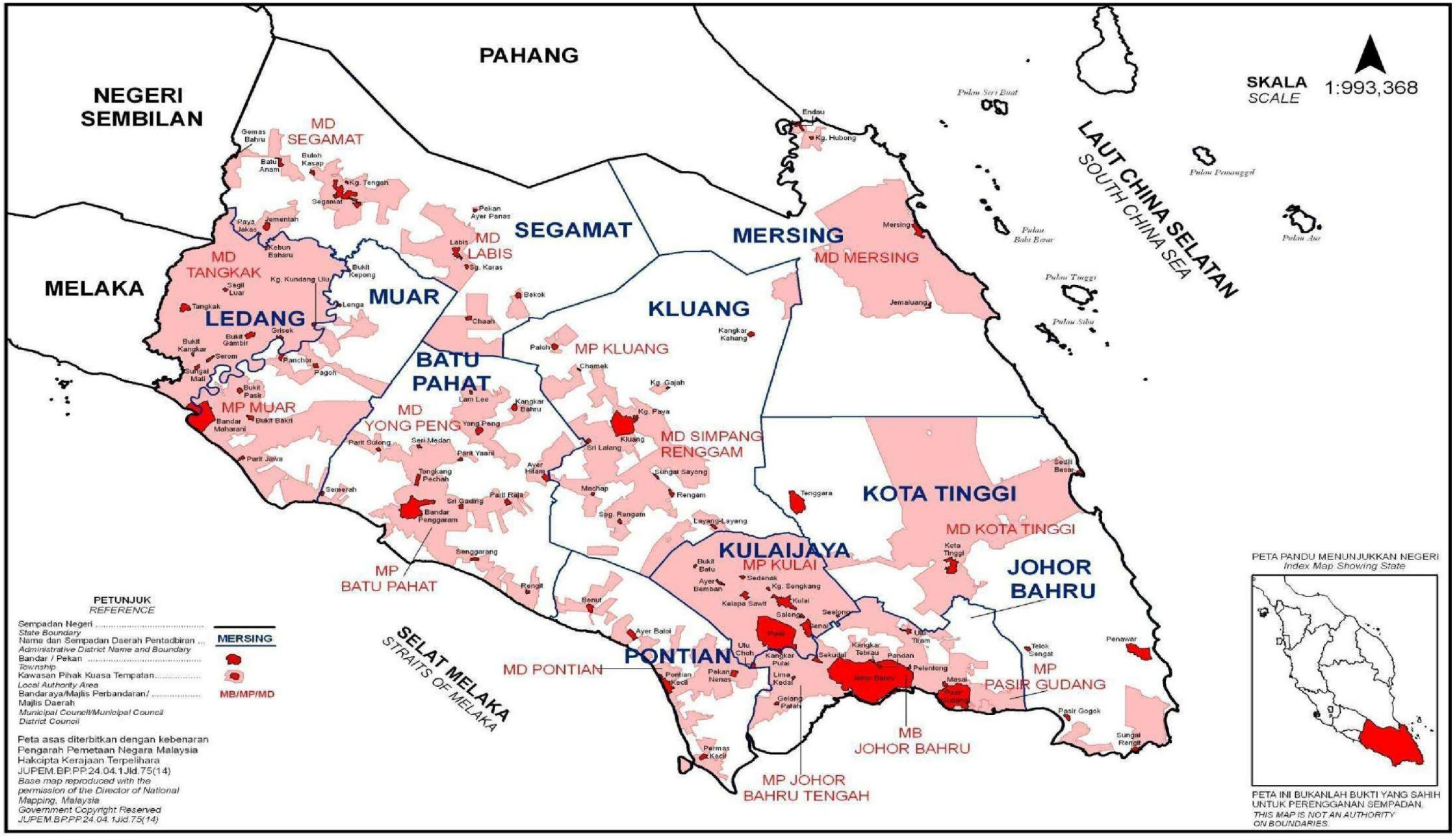


Figure 3. 3: Example of cities' boundaries (dark red) in *Johor State* outlined by the Department of Statistics Malaysia, 2010

The urban-rural boundaries created by both agencies (shown in Figure 3.2 and Figure 3.3) are entirely different from the previous spatial units (e.g., districts or *mukim*). They were created to serve different purposes and define urban areas in varying ways. For example, city boundaries created by the Malaysia Department of Statistics are based on Enumeration Blocks for census purposes; in contrast, city boundaries created by the Malaysian Federal Department of Statistics are based on 'City Profile Reports' and 'National Urbanisation Policy', which are used for the purpose of urban planning and urban development control. As mentioned earlier, the urban-rural boundaries created by these agencies also rely on the definitions used to define urban areas. This has resulted in uncorrelated information between agencies. For example, according to the Malaysia Department of Statistics, there were 149 cities in 2010; on the contrary, the Federal Department of Town and Country Planning in Peninsular Malaysia outlined 288 cities (Federal Department of Town and Country Planning in Peninsular Malaysia, 2016).

The following paragraphs discuss the definition from the Malaysia Department of Statistics because the definition from the Federal Department of Town and Country Planning in Peninsular Malaysia was modified from the former definition.

Urban areas in Malaysia were first defined in 1947 as areas with a population of 1,000 or more. In 1957, the definition was updated to include municipalities, town council areas, town board areas, local council areas, new villages, and villages (Yaakob et al., 2010). In 1970 and 1980, the definition was revised to avoid inclusion of small settlements by increasing the minimum population to 10,000 or more (Hirschman, 1976; Yaakob et al., 2010). The definition was further revised in 1991 and 2000 to include adjoining built-up areas where 60 percent of the population (aged 10 years or more) is engaged in non-agricultural activities and at least 30 percent of housing has modern toilet facilities (Hasan & Nair, 2014). In 2010, the modern toilet facilities criterion was removed, and the minimum age for the working-age group was increased to 15 years or more. These changes were made because almost all houses had modern toilet facilities and the Labour Force Survey showed the working age starts at 15 (Hasan & Nair, 2014). The following statement shows the 2010 definition, which has been used up to the present:

Gazetted areas with their adjoining built-up areas, which had a combined population of 10,000 or more at the time of the Census 2010 or the special development area that can be identified, which at least had a population of 10,000 with at least 60% of the population (aged 15 years and above) were involved in non-agricultural activities.

Malaysia Department of Statistics, 2016

The Federal Department of Town and Country Planning in Peninsular Malaysia (2016) uses a similar definition with extra criteria for defining urban areas: 1) population density of 50-60 people per hectare and 2) the presence of urban infrastructure and facilities. Overall, the agency outlines seven levels of urban hierarchies and the corresponding boundaries (see Figure 3.2): National Growth Conurbation (more than 2.5 million population), Regional Growth Conurbation (1.5 to 2.5 million population), Sub-regional Growth Conurbation (0.5 to 1.5 million population), State Growth Conurbation (0.3 to 0.5 million population), District Growth Conurbation (0.1 to 0.3 million population), Major Settlement Centre (30,000 to 0.3 million population), and Minor Settlement Centre (10,000 to 30,000 population).

The urban-rural boundaries from both agencies do not provide the comprehensive information needed to adopt a differential urbanisation theory approach or to examine urbanisation comprehensively. For example, it is impossible to identify which cities are large, medium, or small from the urban boundaries outlined by the Malaysia Department of Statistics because all cities are simply characterised as urban areas. Further, migration data is recorded simply as urban-rural or rural-urban or urban-urban instead of using more detailed urban and rural classifications. More importantly, there is no information on the socio-economic characteristics needed to analyse the determinants of population change and migration. Although the Federal Department of Town and Country Planning in Peninsular Malaysia provides a detailed hierarchy of cities and their boundaries, other information provided is rudimentary (e.g., total population, population density, and total land area).

Given the limitations, the existing urban-rural units from both agencies are not ideal for applying differential urbanisation theory. In contrast, the existing small-area units (district and *mukim*) provide sufficient information to adopt differential urbanisation theory and analyse urbanisation and population and migration change comprehensively. Therefore, the next section explains the transformation of small-area units (districts and *mukim*) into urban-rural units.

3.3 Developing a new urban-rural classification

Since the existing urban-rural classification cannot be used to examine population change and internal migration comprehensively and to adopt differential urbanisation theory, this section develops a new urban-rural classification for Malaysia using small-area geography (district and *mukim*) due to wider range of data sources. Generally, there is no specific method in classifying urban and rural areas due to different definitions and criteria used across countries. However, common criteria of urban and rural areas include the size of population, population density, and administrative status (Tacoli & Satterthwaite, 2015). Nonetheless, Champion (2003) highlighted that the urban and rural areas must also concern the whole functional units instead of just their pre-defined boundaries. For example, Halliday and Coombes (1995) incorporated more criteria such as built-up areas, hinterland, and the local labour market areas to demarcate functional urban regions in the UK.

The basic requirement to test differential urbanisation theory is to classify each city into large, medium and small. However, there is no guideline mentioned in differential urbanisation theory on how to differentiate the cities. Therefore the classification process is based on available information and author's knowledge about Malaysia's urban environment. This chapter uses the official settlement hierarchy as the main reference to link with differential urbanisation theory in order to create a new settlement hierarchy/type by using district and *mukim* units.

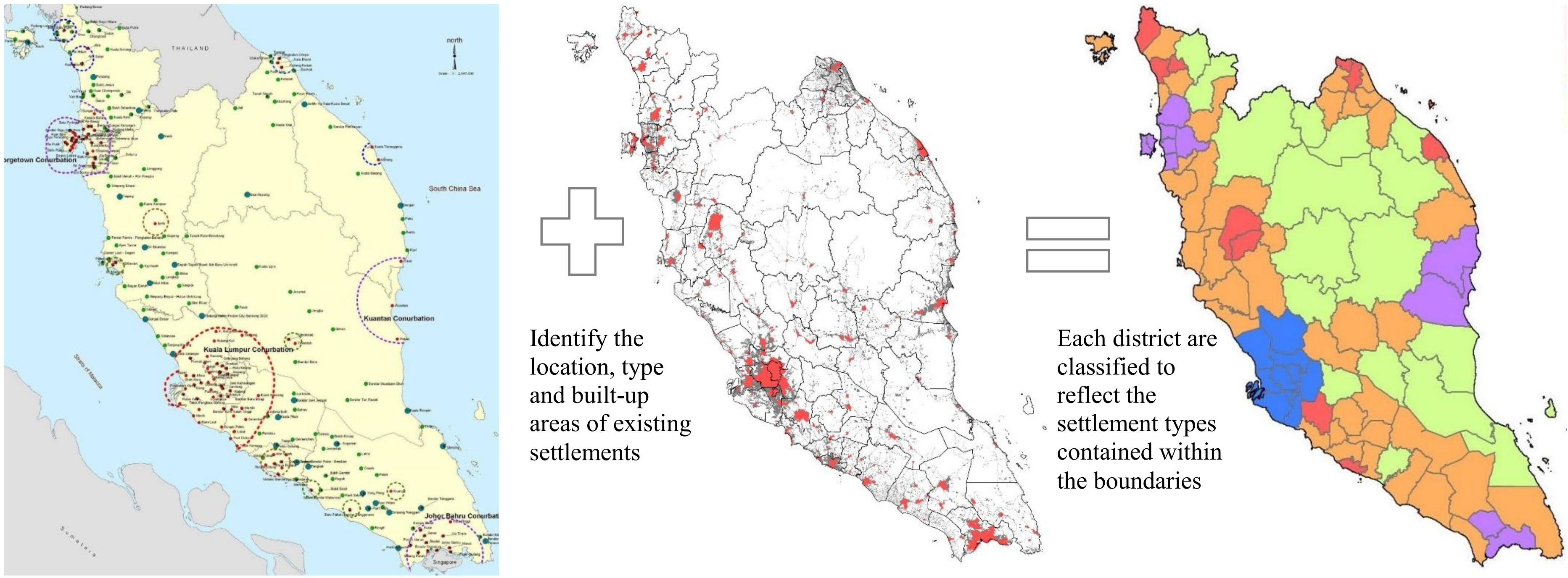
The first step is to aggregate the existing settlement type based on hierarchical levels to reflect the theory assumptions and rename it to fit with the Malaysian context (Table 3.1).

Table 3. 1: Aggregation of existing settlement type to reflect differential urbanisation theory asumption

<table>
<tr><th>No.</th><th>Official settlement hierarchy (Federal Department of Town and Country Planning in Peninsular Malaysia, 2016)</th><th>Settlement hierarchy based on differential urbanisation theory (Geyer & Kontuly, 1993)</th><th>New settlement hierarchy based on differential urbanisation theory in the Malaysian context</th></tr>
<tr><td>1</td><td>National Conurbation</td><td>Primate/Largest city</td><td>Capital metropolitan</td></tr>
<tr><td>2</td><td>Regional Conurbations</td><td rowspan="4">Intermediate-sized cities</td><td>Regional metropolitan</td></tr>
<tr><td>3</td><td>Sub-regional Conurbations</td><td rowspan="3">Intermediate-sized cities</td></tr>
<tr><td>4</td><td>State Conurbations</td></tr>
<tr><td>5</td><td>District Conurbations</td></tr>
<tr><td>6</td><td>Major Settlement Centre</td><td rowspan="2">Small cities</td><td rowspan="2">Small towns/villages</td></tr>
<tr><td>7</td><td>Minor Settlement Centre</td></tr>
<tr><td>8</td><td>Rural Areas</td><td>-</td><td>Remote villages</td></tr>
</table>

Based on Table 3.1, National Conurbation is considered as the primate city since it is known to be the largest settlement type in Malaysia and consist of the capital city, Kuala Lumpur. Regional, Sub-regional, State and District Conurbations are considered as intermediate-sized cities because these cities exist in the regional, sub-regional, state and district context, respectively. Major and minor settlement centres are characterised as small cities because these settlements are local towns including nearby villages. Finally, remote villages are villages located far from the cities.

Once the new settlement hierarchy/type is defined, the next step is to create a new settlement boundary. Following the recommendation by Champion (2003), this chapter considers an extra element (built-up area) to demarcate the new settlement boundary since common criteria (e.g. total population, population density) had already been considered in the existing settlement boundary. Another important reason is because the existing settlement boundary does not reflect the expansion of the built-up areas. For example, the built-up area expands beyond the existing boundary of the National Conurbation. Each district is classified to reflect the settlement types contained within the boundaries. For example, if a district contains part of the National Conurbation (based on location, existing boundaries and built-up areas), then it is classified as Capital metropolitan. For another example, if a district contains many Major or Minor Settlement Centre, it is classified as a small town/village. The overall process is illustrated in Figure 3.4 that shows example transformation of district units into new settlement type in West Malaysia.



Official settlement hierarchy in West Malaysia, (refer Figure 3.2 for the legend)

Official settlement boundary (red) and built-up areas (grey) in West Malaysia, 2010

New settlement type in West Malaysia created using district units

Source: Federal Department of Town and Country Planning in Peninsular Malaysia (2016)

Figure 3. 4: Example transformation of district units into settlement type in West Malaysia

Finally, Table 3.2 shows the number of district and mukim that are aggregated into each settlement type.

Table 3. 2: Aggregation of districts and *mukim* into settlement types

Settlement type		**Number of district(s)**	**Number of *mukim***
Capital metropolitan	Core	1	**8
	Suburban	9	25
	Total	10	33
Regional metropolitan	Core	*0	3
	Suburban	*0	118
	Total	11	121
Intermediate cities		19	115
Small cities/ villages		49	326
Remote villages		55	337
Total District/ Mukim		**144**	**932**

Note:
** The core city and suburban area for regional metropolitan cannot be distinguished because the district unit covers both settlements.*
***The mukim for the core city/ Kuala Lumpur is available only for the most recent census in 2010.*
The list of districts transformed into each settlement type is shown in A1 in Appendices section. Due to the large number of mukim (932), it is not included in this thesis.

The outcome of the aggregation process produces settlement boundaries created by districts and *mukim* units in Figure 3.5 and Figure 3.6, respectively.

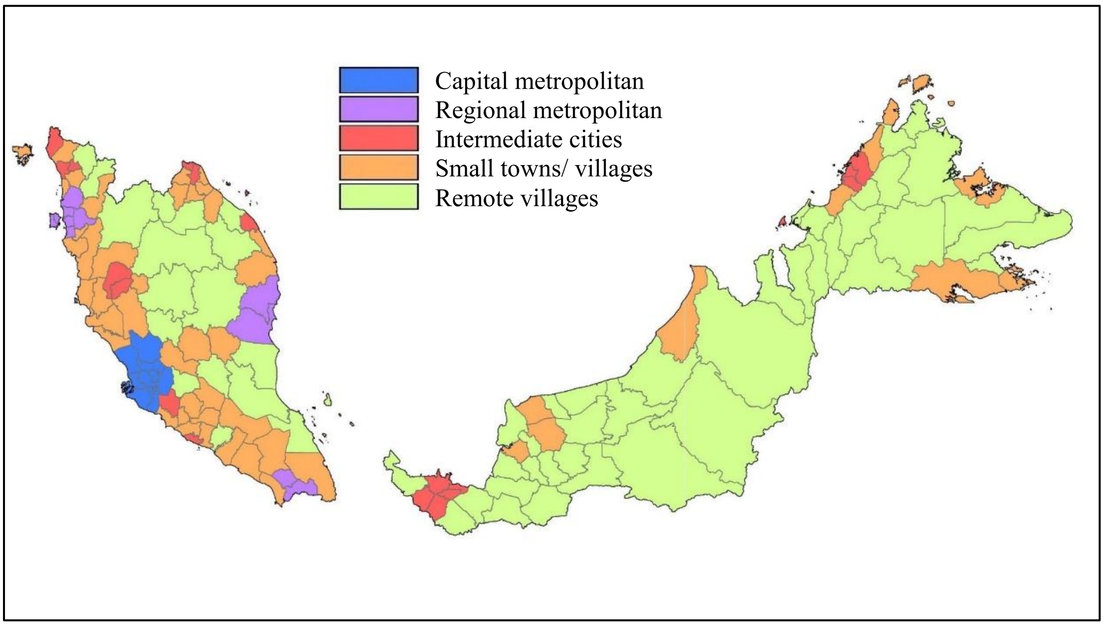


Figure 3. 5: New settlement classification by district units

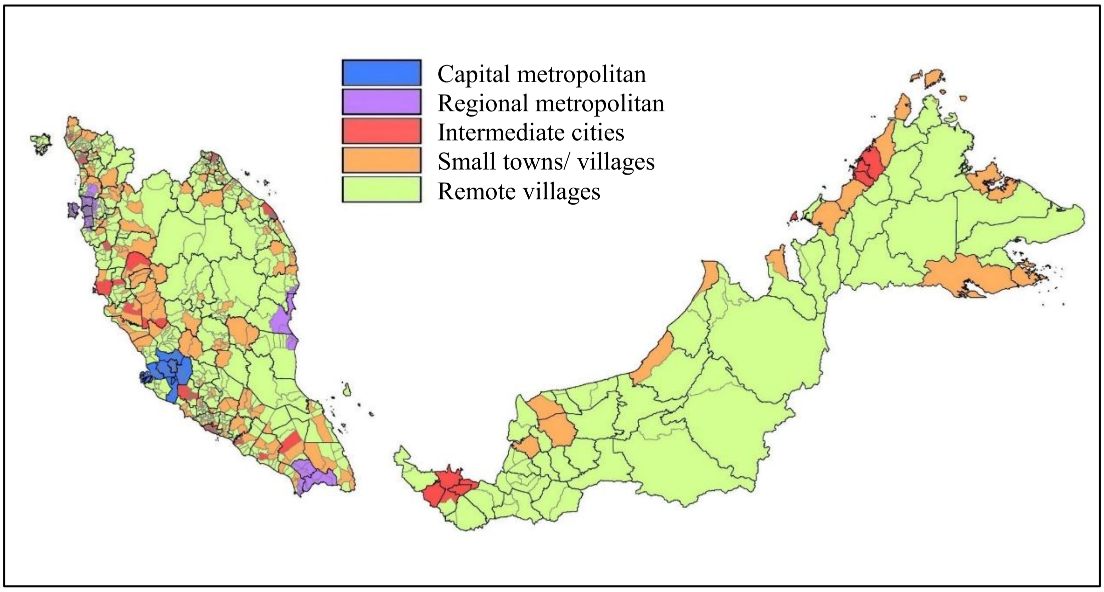


Figure 3. 6: New settlement classification by *mukim* units

However, there are several issues of this chapter's approach. First, unlike mukim units, it is hard to distinguish settlement types using district units because of its larger size than the actual city size. For example, the core city and suburban areas in regional metropolitan areas cannot be distinguished because the district covers both areas. As another example, small towns/villages cannot be separated because of the same issue. Second, the new settlement boundary is used throughout the period studied (1980-2010), where in reality, the boundary changes over time. Finally, because settlement boundary created from district units offer a wider range of data sources, they are used in most chapters (Chapter 5, Chapter 6 and Chapter 7). On the contrary, most analysis in Chapter 4 applies settlement boundary created from *mukim* units since needed data is available. Ideally, the settlement boundary created from *mukim* unit is preferable because they look the closest to the official settlement boundary and the expansion of built-up areas.

These issues are not surprising in geographic studies and are commonly known as Modifiable Area Unit Problem (MAUP). MAUP is a situation when the boundaries of geographical units can be changed which could lead to inconsistent results in statistical analysis (Wong, 2004). According to Dark and Bram (2007), the MAUP consists of two effects: scale effect and aggregation effect. Scale effect happens when there is a variation of results based on the number of areal units used.

Aggregation effect on the other hand happens when the results change because larger number of smaller areal units is aggregated into smaller number of large areal units. The analysis results of this thesis are more likely to receive the latter effect because the districts and *mukim* are aggregated into different settlement type. Furthermore, the results may differ because both boundaries are at different scale and is used throughout the period studied, 1980-2010. This problem is mainly due to the unavailability of detailed data (e.g., migrant's origin and destination, socio-economic characteristics, vital statistics, previous year settlement hierarchy) and previous year geographical boundaries (particularly East Malaysia). However, despite of the limitations and MAUP, surprisingly, the comparisons of analysis that apply district and mukim boundaries in Chapter 4 produce more or less similar results (See Section 4.5.1 for more detail).

3.4 Conclusion

The main objective of this chapter is to explain the process of creating a new urban-rural classification to fit with the differential urbanisation theory approach. This chapter begins by introducing existing spatial units (national, state, district, *mukim*, and urban-rural) in Malaysia to provide a clear understanding of the country's geographical background. Generally, different countries use different definitions and measurements to define urban and rural areas. Malaysia is no exception because urban and rural areas are defined and measured by multiple government agencies. This situation has led to a lack of correlation of the information from different agencies (e.g., number of cities, total population, and population density). Furthermore, only basic information is provided for existing urban-rural units (e.g., total population, population by gender, age group and ethnicity, and total migration by urban and rural areas) and hence the author could not use this to adopt differential urbanisation theory and comprehensively examine urbanisation. In contrast, detailed information is available for other small-area units such as the district and *mukim*. This chapter integrates existing information from different agencies to build new urban and rural classifications and spatial units using districts and *mukim*. The new classifications and spatial units are used in the subsequent chapters to examine population growth, internal migration, and urbanisation in Malaysia.

Chapter 4

Population growth and urbanisation in Malaysia, 1980-2010 – a macro and micro perspective

4.1 Introduction

Urbanisation and urban population growth are closely related whereby urbanisation results from an increase in the proportion of the population living in urban areas. Previous studies that have adopted differential urbanisation theory used population growth as the main indicator to examine urbanisation and the transition process (see Gedik, 2003; Gwebu, 2006; Mookherjee, 2003). Building on this, this chapter examines urbanisation by observing population growth in urban and rural areas in Malaysia from 1980-2010 using the spatial units developed in Chapter 3. Before applying the differential urbanisation theory approach, this chapter first focuses on population change in terms of the existing spatial units in Malaysia at the macro level (states and regions) in Section 4.3.1 and the micro level (districts and *mukim*) in Section 4.3.2. Due to the only large number of the districts and *mukim*, Section 4.3.2 observes population change for only a few that are part of metropolitan regions. These regions were formed by the Malaysian government in 2006to create a balance in regional development growth. They represent the highest level of the settlement hierarchy and have seen significant changes, which have been highlighted in previous studies (see Abdullah, 2003, 2012; Abdullah & Mohd, 2009; Osman et al., 2017, 2009). However, a shortcoming of Section 4.3.2 is that it only observes population change in some cities. Therefore, in Section 4.4, differential urbanisation theory is applied to tackle this gap by observing population change for all settlements. Expanding on Section 4.4, Section 4.5 examines demographic (gender, age group, and ethnicity) and socio-economic (education attainment, occupation type, and industry type) changes by settlement type. Overall, Malaysia may have experienced the final stage of urbanisation, Advanced Primate City (APC), due to rapid population growth in the capital metropolitan core and suburban areas and from the shrinking dominance of the capital metropolitan core since 1980. However, Chapter 5 puts the theory to test even further by observing the concentration and deconcentration of migrants between settlements.

4.2 Data and Methods

4.2.1 Data

Table 4. 1: Datasets used in Chapter 4

Source	Scale	Types of information	Years available
Department of Statistics Malaysia	State, District *Mukim*	Total population by: age group	1980 1991 2000 2010
IPUMS International	State District	Total population by: gender age group ethnicity education attainment types of working industries types of occupation	1980 1991 2000

Based on Table 4.1, this chapter uses a combination of datasets obtained from the Department of Statistics Malaysia and IPUMS International. Note that IPUMS International dataset is also considered official because they retrieved it from the Department of Statistics Malaysia. However, the sample size of IPUMS International dataset only covers 2 percent from the overall population, smallest geography is district level and available from 1980 until 2000. On the contrary, Department of Statistics Malaysia dataset covers the overall Malaysian population, smallest geography is *mukim* level and available from 1980 until 2010.

The main reason IPUMS International dataset is needed in this chapter is due to the limited small-area units (district/*mukim*) information provided by the Department of Statistics Malaysia. Most information on demographic and socio-economic characteristics is obtained from IPUMS International which is important to examine population change and urbanisation, as well as to test differential urbanisation theory comprehensively. Many studies that adopted differential urbanisation theory only focussed on the overall population and/or net migration change in their analysis (Bonifazi & Heins, 2003; Gedik, 2003; Gwebu, 2006; Mookherjee, 2003; Tammaru et al., 2004) instead of considering demographic and socio-economic perspective which could lead to better understanding in explaining urbanisation process. Different areas generally consist of different characteristics of population. For example, more people work as professionals and semi-professionals in cities, more farming and agricultural jobs in rural areas, higher wages in cities than in rural and more educated people in cities.

There is also growing evidence of contradicting population patterns among countries. Countries like United States, Britain, France, Australia, Italy, and West Germany experienced a major population shifts towards non-metropolitan regions where there was a large number of elderly moved and lived in rural areas (Argent & Rolley, 2008; Berry, 1980; Champion, 2003; Coombes et al., 1989; Fielding, 1982; Kontuly & Vogelsang, 1988). On the contrary, more elderly in South Korea prefer to live in urban areas than rural areas upon retirement due to the high accessibility to health services and living amenities despite many policies imposed by the government to encourage them to out-migrate to rural areas (Kim & Han, 2014). Situation in South Korea is similar in developing and less developed countries (most likely in Malaysia too), with governments implementing various policies (e.g., resettlement schemes, industrial decentralisation, and nativist systems) to discourage people living in urban areas (Lall et al., 2006). Therefore, the use of IPUMS International dataset is ideal in this thesis in order to understand the demographic and socio-economic perspective in relation to urbanisation process in Malaysia.

4.2.2 Methods

Once the datasets have been identified, the next step is to apply methods for the the analysis. Section 4.3 applied basic demographic techniques (population share and growth) to examine population distribution and change at all spatial scales in Malaysia. However, due to their large numbers of the small-area units (144 for District and 933 for *Mukim*), it is impossible to examine it all. Therefore, specific areas are selected – the metropolitan regions. Metropolitan region is known as the largest cities in terms of settlement hierarchy in Malaysia. There are four metropolitan regions starting with the largest: *Kuala Lumpur*, *Georgetown, Johor Bahru* and *Kuantan.* These regions are known to have significant population change based on previous literature (See Abdullah, 2003, 2012; Abdullah & Mohd, 2009; Osman et al., 2009). The equations that are used in Section 4.3 are as follow:

For the macro level (Section 4.3.1):

$$\%\,Population\ Share = \frac{Population\ \ of\ a\ State}{Population\ of\ Malaysia} \times 100$$

$$\%\,Population\ Growth = \frac{(Present\ population - Past\ population)}{Past\ population} \times 100$$

For the micro level (Section 4.3.2), the population share and growth are calculated for each metropolitan regions:

$$\%\,Population\ Share = \frac{Population\ \ of\, a\ District/Mukim}{Population\ of\ Metropolitan\ region} \times 100$$

Because Section 4.3 only observes population growth for the metropolitan region, Section 4.4 tries to fill the gap by examining all settlements growth which can be achieved through the application of differential urbanisation theory by using the new settlement classification built in Chapter 3. Similar to Section 4.3, the same calculations will be used in Section 4.4. Furthermore, two additional equations are also applied to examine the fast and slow pace of population growth in comparison to the starting year and also to Malaysia's overall growth:

Change relative to 1980 population:

$$R_{S}^{1980+n} = \frac{P_{S}^{1980+n}}{P_{S}^{1980}}$$

Where S is settlement type, R^{1980+n} is the change of subsequent year population to 1980 population, P^{1980+n} is population of the subsequent year, and P^{1980} is 1980 population.

Change relative to Malaysia population:

$$R_{M}^{1980+n} = \frac{P_{M}^{1980+n}}{P_{M}^{1980}}$$

$$R_{SM}^{1980+n} = \frac{R_{S}^{1980+n}}{R_{M}^{1980+n}}$$

Where *M* is Malaysia as a whole, *SM* is the change of settlement type relative to Malaysia as a whole, and the rest is the same as in (1).

Once the overall change and pattern are observed, Section 4.5 attempts to observe population characteristics in each settlement type: by gender, age groups, ethnicity, education attainment, types of working industries and types of occupation. As mentioned earlier, due to the limitation of data from the Department of Statistics Malaysia, Section 4.5 uses data provided by IPUMS International and since the dataset only covers 2 percent from the overall population, therefore the values are then converted into percentages.

4.3 Population growth in existing spatial units

Before implementing the differential urbanisation theory approach, this section observes population growth of the existing spatial units in Malaysia (e.g., states, districts, and *mukim*). Although this thesis is focused on the micro level and small-area units, it is important to include a macro-level perspective to gain a general idea of population change on a larger scale and how it relates to small-area change. In the Malaysian context, regions and states are considered large-area units while districts and *mukim* are considered small-area units. Section 4.3.1 discusses population change at the regional and state levels, followed by Section 4.3.2, which discusses small-area change (district and *mukim*).

4.3.1 Macro units

To better understand large-area change, the states in Malaysia are grouped into two main regions: West Malaysia and East Malaysia. West Malaysia consists of four sub-regions (Central, Northern, Southern, and East Coast) while East Malaysia has no sub-regions (Figure 4.1). Table 4.2 outlines these regions.

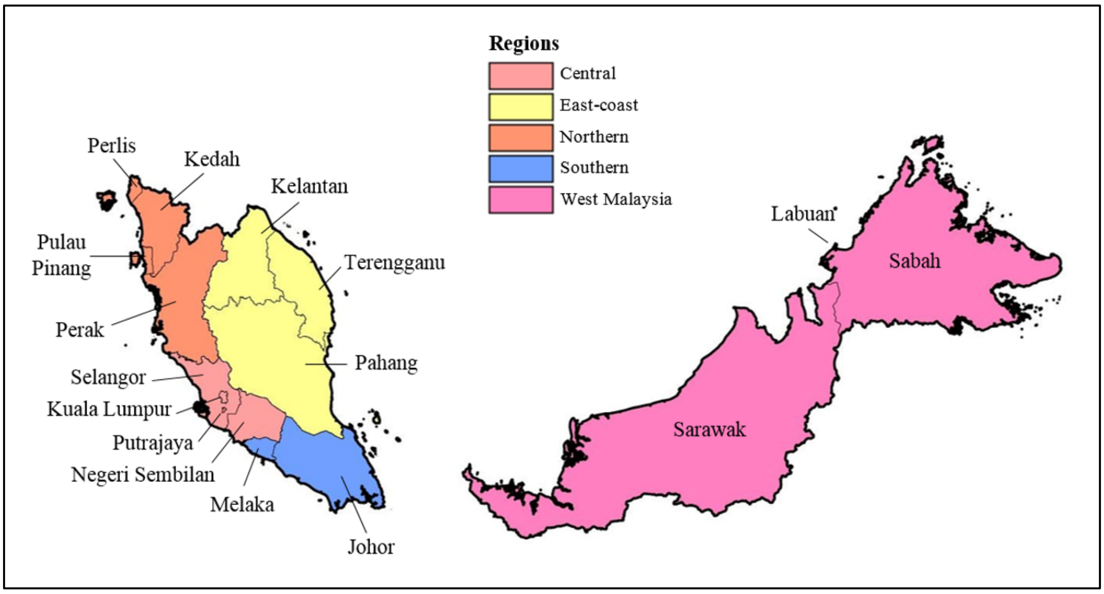


Figure 4. 1: States and regions in Malaysia

Table 4. 2: Population share and growth by region and state in Malaysia, 1980-2010

Region		State	Share (%)				Growth (%)		
			1980	1991	2000	2010	80-91	91-00	00-10
West Malaysia	Central	*F.T. Kuala Lumpur*	7	6.5	5.9	5.8	24.5	14	21.7
		*F.T. Putrajaya**	0	0	0	0.2	0	0	0
		Selangor	10.9	13.1	17.8	19.4	61.1	72.1	35.2
		Negeri Sembilan	4.2	3.9	3.7	3.6	25.7	19.8	18.9
	Northern	*Kedah*	8.2	7.4	7.1	6.9	20.8	20.6	20.9
		Perak	13.3	10.7	8.9	8.4	7.7	5.1	16.5
		Perlis	1.1	1	0.9	0.8	27	7.9	13.8
		Pulau Pinang	6.9	6.1	5.5	5.6	18.1	15.7	24
	Southern	*Johor*	12	11.8	11.6	11.8	31	24.9	25
		Melaka	3.4	2.9	2.7	2.9	13.3	19.5	30.5
	East coast	*Kelantan*	6.5	6.7	5.8	5.4	37.5	9	14.2
		Pahang	5.9	5.9	5.5	5.2	35.9	17.6	17.2
		Terengganu	4	4.4	4	3.7	45.9	14.9	14.9
East Malaysia		*Sabah*	7.3	9.9	11.1	11.3	81.5	42.3	26.3
		Sarawak	9.4	9.4	9.1	8.7	33	22.3	19.4
		F.T. Labuan	0	0.3	0.3	0.3	0	30.7	18.4

*Note: *The reason Putrajaya share and growth was 0 is that it was part of Selangor before it was separated into a Federal Territory in 2001. Labuan was part of Sabah and became a Federal Territory in 1984.*

Central Region – This region includes the national capital (*F.T. Kuala Lumpur*) and the national administrative centre (*F.T. Putrajaya*) of Malaysia. The result shows the population share in *Kuala Lumpur* continuously declined throughout the entire period. In contrast, *Selangor* had a rapid increase in its population share, from 10.9 percent to 17.8 percent. The contradictory patterns between these states show *Selangor* has had a higher population concentration than *Kuala Lumpur* since 1980. This is due to many reasons. First, *Kuala Lumpur* is confined by limited space and is surrounded by *Selangor* and thus there is no potential for urban expansion beyond its administrative borders. Second, since *Kuala Lumpur* is fully urbanised (100 percent urban population) and has no potential of expanding, the 'spill-over effect' of urban development to *Selangor* has led to the establishment of many new townships and major infrastructure (e.g., highways, and an international airport) (Hasan & Nair, 2014). Third, *Selangor* has received huge economic investments due to its strategic location. It surrounds the national capital and largest city (*Kuala Lumpur*) as well as the national administrative centre (*Putrajaya*) and is the location of one of the largest seaports in Malaysia (*Port Klang*). Further, there has been a mushrooming of multinational companies in *Selangor*.

Northern Region – The population share for all states in this region declined throughout the period. *Perak* experienced the largest decline, from 13.3 percent to 8.4 percent, due to its location next to *Selangor*, which attracted more people. Similar to the experience of *Kuala Lumpur*, *Pulau Pinang* is known as one of the most developed and urbanised states but its population share decreased. There is no clear explanation for this; hence, more observation at the district and *mukim* levels is required.

Southern Region – *Johor* has the largest population share in this region. The strong population share in *Johor* is mainly due to its strategic location neighbouring another country, *Singapore*. It has one of the largest and busiest seaports in Malaysia (i.e., *Johor Port* and *Port of Tanjung Pelepas*). Furthermore, it also includes the third largest city in Malaysia, *Johor Bahru City*.

East Coast Region and East Malaysia – The population share in East Coast Region states (*Kelantan, Pahang,* and *Terengganu*) showed no major change, with a slow declined during the study period. Finally, the population in East Malaysia (*Sabah, Sarawak,* and *Labuan*) saw strong and stable growth, likely due to the natural increase of the population in rural areas. Generally, these states are not as developed and urbanised as states in West Malaysia, with most people residing in rural areas.

Overall, imbalanced population growth in these regions can be seen from 1991 due to the dominance of states in the Central Region, particularly in *Selangor*. Increasing migration to West Malaysian states has created problems such as imbalanced urban development and urban poverty in main cities (Masron et al., 2012). Furthermore, rapid development in Central Region states is evident from the large increase in the number of townships and major infrastructure development (i.e., airports, highways, and light rail transit). Moreover, the distribution of urban centres is spatially uneven, with more cities located in high-density areas in West Malaysia than in East Malaysia. These urban centres have existed since the colonial period and have grown significantly over the years. Due to imbalanced development between the regions, the Malaysian government included initiatives aimed at balanced regional growth in the Ninth and Tenth Malaysia Development Plan (Hasan & Nair, 2014). In line with such policies, several mega-urban regions have been formed, which are discussed in the next section.

4.3.2 Micro-units

This section examines population change and growth from a micro perspective considering district and *mukim* units. Due to the large number of these units, it is impossible to examine all of them. The selection of districts and *mukim* is based on the metropolitan regions formed by the Malaysian government in 2006 to create balanced regional development (Hasan & Nair, 2014). These regions represent the highest level of the settlement hierarchy and have seen significant changes, which have been highlighted in previous studies (see Abdullah, 2003, 2012; Abdullah & Mohd, 2009; Osman et al., 2017, 2009). These metropolitan regions comprise cities that were merged through physical and economic expansion (Figure 4.2).

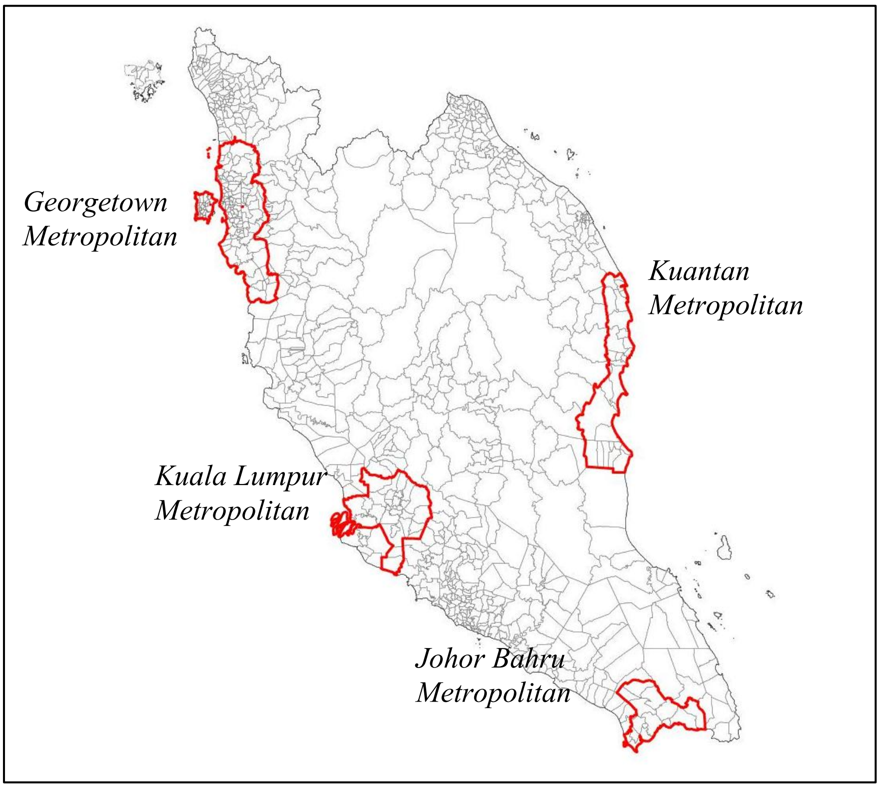


Figure 4. 2: Official metropolitan regions in Peninsular Malaysia

Figure 4.2 shows the official metropolitan regions that were formed by the Federal Department of Town and Country Planning in Peninsular Malaysia. There are four metropolitan regions: *Kuala Lumpur, Georgetown, Johor Bahru,* and *Kuantan*. *Kuala Lumpur* is the largest metropolitan region and is a combination of eight districts and 24 *mukim*; *Georgetown* consists of 13 districts and 137 *mukim*; *Johor Bahru* comprises of three districts and 12 *mukim*; and, *Kuantan* contains four districts and 28 *mukim*.

The results of population share and growth by district and *mukim* for each metropolitan region are presented in Table 4.3 and Table 4.4, respectively.

Table 4. 3: Population share and growth by district in metropolitan cities, 1980-2010

Metropolitan region	District	Share (%)				Growth (%)		
		1980	1991	2000	2010	80-91	91-00	00-10
Kuala Lumpur	*Kuala Lumpur*	47.2	38.1	28.1	25.3	24.5	14	21.7
	Putrajaya	0	0	0	9.1	0	0	0
	Gombak	8.5	11.7	11.6	10.6	112.4	52.4	24.4
	Klang	14.3	13.5	13.9	13.4	45.7	58.1	30.9
	Petaling	18.5	21.1	25.5	28.1	75.9	87	49.1
	Sepang	2.4	1.8	2.3	3.3	18.8	98.7	90.9
	Ulu Langat	9.1	13.8	18.6	18.1	132.7	108.9	31.7
Georgetown	*Barat Daya*	5	6.8	7.4	7.3	60.7	29.6	23.9
	S.P. Selatan	4.6	4.7	5.5	6.2	18.5	38.9	41.6
	S.P. Tengah	10.5	13	13.7	13.4	45.9	24.5	23.4
	S.P. Utara	12.8	12.4	11.4	10.7	13.5	8.6	18.3
	Timur Laut	25.3	21.8	19.4	18.9	1.3	5.2	22.7
	Bandar Baharu	2.1	1.9	1.8	1.5	6.5	12.4	8.9
	Kuala Muda	12.5	14	15.8	16.4	32.3	33.6	30.5
	Kulim	5.5	6.7	8.6	10.2	42.8	51.8	48.8
	Yan	1.3	1	1	0.9	-1.9	12.2	15
	Kerian	10.1	8.2	7.1	6.5	-4.4	2.8	15.7
	Larut dan Matang	10.4	9.7	8.4	8	10.4	2	21.1
Johor Bahru	*Johor Bahru*	76.9	81.2	82.6	82.3	78.6	54.7	44.6
	Kulaijaya	16.4	14.7	14.3	15.1	51.6	47.3	53.9
	Pontian	6.7	4.1	3.1	2.5	1.4	20.1	23.9
Kuantan	*Kuantan*	54.1	52.1	54.1	55.9	49.7	35.8	31.7
	Pekan	11.4	9.2	8.5	7.8	24.3	21.8	17
	Dungun	16	17.9	17.7	16.6	73.9	28.7	19.6
	Kemaman	18.4	20.8	19.7	19.7	75.3	24.1	27.9

*Note: *The reason Putrajaya share and growth was 0 is that it was part of Selangor before it was separated into a Federal Territory in 2001.*

Table 4. 4: Population share and growth by *mukim* in metropolitan cities, 1980-2010

Metropolitan region	District	*Mukim*	Share (%) 1980	1991	2000	2010	Growth (%) 1980-91	1991-00	2000-10
Kuala Lumpur	***Kuala Lumpur***	***Kuala Lumpur***	**47.2**	**38.1**	**28.1**	**25.3**	**24.5**	**14**	**21.7**
	Putrajaya	*Putrajaya*	0	0	0	1.1	0	0	0
	Gombak	*Batu*	3.9	5.9	5.4	4.5	132.2	41.3	13.3
		Rawang	1.8	1.7	2.6	3.2	50.6	133.5	64.4
		Setapak	1.1	1.6	1.5	1.2	129.5	40	11.1
		Ulu Kelang	1.7	2.5	2.1	1.7	119.4	30.5	12.4
	Klang	*Bandar Klang*	1.5	0.8	0.4	0.2	-21.9	-18.3	-44
		Kapar	4.4	4.4	4.7	4.2	55.7	62.6	21.5
		Klang	8.4	8.3	8.8	9.1	52.5	62.6	39.3
	Petaling	*Bandar Petaling Jaya*	3.8	2.3	1.4	1	-8.3	-0.8	-8.8
		Bukit Raja	0.4	0.8	0.8	1.9	246.8	45.8	230.4
		Damansara	4.8	7.8	9.8	8.2	148.5	93.9	13.9
		Petaling	4	4.9	7.7	9.6	87.4	144	68.9
		Sungai Buloh	5.5	5.3	5.8	7.4	50.4	68.5	72.4
	Sepang	*Dengkil*	1.2	1	1.6	2.7	37	133.1	135.2
		Labu	0.3	0.2	0.2	0.1	3.5	42.4	20.3
		Sepang	0.9	0.6	0.6	0.4	0.4	56.4	-2.6
	Ulu Langat	*Ampang*	3.6	6.5	7.7	5.5	176.7	83.4	-4.3
		Beranang	0.4	0.3	0.3	0.8	23	50.5	253.7
		Cheras	1.1	1.8	3.5	3.9	150.1	194.4	49.5
		Kajang	2.4	3.6	4.9	5.5	134.2	111.5	49.2
		Semenyih	0.8	0.8	1.1	1.6	46.5	115.4	103.1
		Ulu Langat	0.7	0.7	1	0.9	44.3	130.5	18.1
		Ulu Semenyih	0.1	0.1	0.1	0.1	41.1	56.3	5.9
Georgetown	*Barat Daya*	*Mukim 1*	0.1	0.2	0.2	0.2	81.4	16.3	10.1
		Mukim 2	0.2	0.2	0.2	0.1	-8	-3.4	-18.9
		Mukim 3	0.1	0.1	0.1	0.1	24.2	6.6	-31.3
		Mukim 4	0.2	0.1	0.1	0.1	-0.7	-23	24.1
		Mukim 5	0	0	0	0	-40.8	-57.5	24.1
		Mukim 6	0.2	0.3	0.3	0.3	42.6	30	26.1
		Mukim 7	0.1	0.1	0.1	0.1	2.8	-8.8	-12.9
		Mukim 8	0.1	0.1	0	0	-0.1	-17.7	41.2
		Mukim 9	0.3	0.4	0.5	0.6	38.2	51.5	34.8
		Mukim 10	0.1	0.1	0.1	0.1	5.8	49.5	21.8
		Mukim 11	0.3	0.4	0.5	0.6	49.1	71.6	33.9
		Mukim 12	2.3	3.9	4.4	4.5	102.4	32.2	29.5
		Mukim A	0.2	0.1	0.1	0.1	7.2	-0.9	-32.8
		Mukim B	0.1	0	0.1	0.1	-5.7	77.5	18.3
		Mukim C	0.1	0.1	0.1	0.1	9.9	39.5	30.1
		Mukim D	0.1	0.1	0.1	0.1	13.2	4.8	-21.8
		Mukim E	0.1	0.1	0.1	0.1	25.3	8.3	-15.3
		Mukim F	0.1	0.1	0.1	0.1	55.8	-7.4	42.3
		Mukim G	0	0	0.1	0.1	14.4	158.7	37.8
		Mukim H	0.1	0.1	0.1	0	35.8	4	-11.3
		Mukim I	0.1	0.1	0.1	0	34.7	10.7	1.7
		Mukim J	0.1	0.1	0.1	0	26.6	10.5	-21.6
	S.P. Selatan	*Mukim 1*	0.2	0.2	0.2	0.1	28.6	9.2	-5.9
		Mukim 2	0.1	0.1	0.1	0.1	6.4	63.5	1.5
		Mukim 3	0	0	0.1	0.1	-10.5	366.3	-4.5
		Mukim 4	0.2	0.2	0.2	0.3	-4.2	43.2	83.8
		Mukim 5	0.3	0.3	0.3	0.3	7.1	30.8	34.2
		Mukim 6	0	0	0	0.1	-12.4	-41.1	1,213.10

	Mukim 7	0.3	0.4	0.5	0.7	41.5	48.6	82.3
	Mukim 8	0.2	0.1	0.1	0.1	3.1	-1.9	10
	Mukim 9	0.5	0.6	0.6	0.8	42.6	31.5	57.1
	Mukim 10	0.5	0.5	0.4	0.4	8.7	9.9	1.1
	Mukim 11	1.2	1.2	1.2	0.9	19	13.9	-1.2
	Mukim 12	0.2	0.2	0.2	0.1	28.3	2.3	-30.2
	Mukim 13	0.1	0.1	0.1	0.2	3.5	16.6	164
	Mukim 14	0.3	0.4	0.6	0.9	20.2	115.1	79.5
	Mukim 15	0.5	0.5	0.9	1.1	14.6	102.8	55.1
	Mukim 16	0	0	0	0	47.7	1.4	35.3
S.P. Tengah	*Mukim 1*	1.2	2.6	2	1.6	142.9	-8.2	-1.3
	Mukim 1A	0.6	0.5	0.6	0.5	12.8	27.5	14.1
	Mukim 2	0.3	0.3	0.3	0.2	17.4	10.8	-6.2
	Mukim 3	0.3	0.3	0.3	0.3	28.6	38.1	-1.9
	Mukim 4	0.5	0.4	0.4	0.4	6.5	18.3	18
	Mukim 5	0.2	0.2	0.2	0.2	16.5	13.8	-1.8
	Mukim 6	0.4	0.7	1	1.9	124.7	69.9	127.4
	Mukim 7	0.1	0.2	0.1	0.2	32.4	15.4	61
	Mukim 8	0.7	0.6	0.4	0.3	-5.4	-23.1	-4.3
	Mukim 9	0.7	0.7	0.5	0.4	6.7	-2.5	-5.6
	Mukim 10	1.5	1.1	0.8	0.4	-12.2	-12.3	-37.9
	Mukim 11	0.6	1.3	1.5	1.2	182.5	30.9	2
	Mukim 12	0.2	0.2	0.2	0.3	28.9	17.4	48.8
	Mukim 13	0.2	0.2	0.4	0.7	7.1	171.8	99.3
	Mukim 14	0.5	0.7	1.2	1.3	66.7	99.5	38
	Mukim 15	1	1.5	1.9	2.1	74	50.8	41
	Mukim 16	0.4	0.4	0.5	0.6	12.4	40.4	53.9
	Mukim 17	0.2	0.1	0.1	0.1	-13.2	-11.1	39.5
	Mukim 18	0.1	0.1	0.1	0.1	24.1	21.5	44.4
	Mukim 19	0.2	0.2	0.2	0.1	28.3	23.9	-2.5
	Mukim 20	0.5	0.6	0.7	0.5	28.7	46.5	-12.8
	Mukim 21	0.2	0.2	0.2	0.2	16.6	27.8	4.4
S.P. Utara	*Mukim 1*	0.4	0.4	0.4	0.3	20.9	17.6	2.5
	Mukim 2	0.4	0.3	0.3	0.2	7.5	10.1	-1.7
	Mukim 3	0.5	0.5	0.5	0.4	15.5	13.4	-4.3
	Mukim 4	0.4	0.4	0.3	0.4	18.6	9.1	61.2
	Mukim 5	0.4	0.4	0.4	0.3	13	9	1.2
	Mukim 6	0.8	0.8	0.8	1.4	19.4	13.7	113.8
	Mukim 7	0.7	0.7	0.6	0.7	5.9	6.8	38.5
	Mukim 8	0.5	0.5	0.6	0.6	34.6	39.9	19
	Mukim 9	0.7	0.8	0.7	0.4	37.9	0.5	-21
	Mukim 10	0.2	0.2	0.2	0.2	19.1	-2.4	4
	Mukim 11	0.5	0.6	0.8	0.7	35.2	67.1	10.3
	Mukim 12	1	1	1.1	1	20.3	33.2	6.5
	Mukim 13	0.3	0.3	0.3	0.3	27	7.2	17.9
	Mukim 14	3.5	3.4	2.7	2.5	13.1	-5.5	14.7
	Mukim 15	2.2	1.6	1.2	0.9	-13.2	-11.2	-4.9
	Mukim 16	0.3	0.3	0.3	0.4	33.9	17.1	60.2
Timur Laut	***Bandaraya Georgetown***	**15.4**	**11.7**	**8.4**	**7.3**	**-10.8**	**-14.7**	**9.3**
	Mukim 13	4.7	6	8.2	8.6	49.2	60.5	32.7
	Mukim 14	0.2	0.1	0.1	0.1	-38.6	5.8	-6.5
	Mukim 15	0	0	0	0	-95.3	92.6	-92.3
	Mukim 16	2.2	1.5	0.9	0.6	-20.3	-32.8	-12.5
	Mukim 17	0.2	0.4	0.2	0.6	81.8	-29.1	237.5
	Mukim 18	2.4	2.1	1.6	1.7	0.9	-8.8	34
Bandar Baharu	*Bagan Samak*	0.8	0.7	0.6	0.5	9.3	-3.3	19.9
	Kuala Selama	0.2	0.2	0.2	0.2	-6.7	25	6.1
	Permatang Pasir	0	0	0	0	3.3	3.5	1.7
	Relau	0.1	0.1	0.1	0.1	10.2	34.6	32.4
	Serdang	0.6	0.5	0.5	0.4	10.5	19.7	-1.9
	Sungai Batu	0.2	0.2	0.2	0.1	-1.4	11.7	-5.9
	Sungai Kechil	0.1	0.1	0.1	0.1	9	37.2	6.3

	Kuala Muda	*Bujang*	0.3	0.3	0.3	0.5	13.5	23.8	82.4
		Bukit Meriam	0.3	0.3	0.3	0.2	4.8	15.8	-5.8
		Gurun	1.8	1.7	1.6	1.5	15.6	8.8	14.4
		Haji Kudong	0.1	0.1	0.1	0	-3.6	4.2	-10
		Kota	0.2	0.2	0.2	0.1	12.4	9.5	-9
		Kuala	0.2	0.1	0.1	0.1	12	21.7	-3.6
		Merbok	0.8	0.7	0.6	0.5	-0.7	10.9	-1.5
		Pekula	0.4	0.5	0.6	0.6	37.3	43.5	22.1
		Pinang Tunggal	0.2	0.2	0.4	0.6	-9.6	124.4	109.5
		Rantau Panjang	0.2	0.1	0.1	0.1	4.7	3	2.2
		Semeling	0.9	0.8	0.7	0.8	1.5	9.5	48.5
		Sidam Kiri	0.5	0.4	0.3	0.3	-4.6	5	11
		Simpor	0.3	0.3	0.3	0.2	14	17.9	6
		Sungai Pasir	1.3	2.4	3.1	3	114.5	51.1	20.1
		Sungai Petani	4.5	5.5	6.7	7.5	44.2	44.4	41.4
		Telui Kiri	0.6	0.5	0.5	0.4	-1.6	11.1	11.7
	Kulim	*Bagan Sena*	0.3	0.3	0.3	0.2	3.1	15.4	8.9
		Junjong	0.2	0.2	0.2	0.2	10	7.6	8.2
		Karangan	0.3	0.4	0.4	0.3	53.5	14.1	9.6
		Keladi	0.2	0.8	1.2	1.2	299.1	79	27.7
		Kulim	1.4	1.7	2.4	2.3	43.6	66.8	24.2
		Lunas	0.5	0.6	0.7	1	53.3	36.1	77.5
		Naga Lilit	0.4	0.4	0.4	0.9	20.3	27.2	171.4
		Padang China	0.5	0.4	0.4	1	-2.7	21.7	194.3
		Padang Meha	0.5	0.4	0.3	0.3	-4.1	-6.2	35.4
		Sidam Kanan	0.5	0.5	0.6	0.9	23.6	29.7	97
		Sungai Seluang	0.3	0.6	1.3	1.2	160.8	173.2	14.2
		Sungai Ular	0.2	0.2	0.2	0.4	24.7	51.1	118.5
		Terap	0.3	0.2	0.2	0.1	8.8	-6.2	-1.8
	Yan	*Singkir*	0.2	0.2	0.1	0.1	-0.3	3.3	0.5
		Yan	1.1	0.9	0.9	0.8	-2.2	13.9	17.4
	Kerian	*Bagan Serai*	2.3	1.8	1.6	1.6	-8.1	2.9	27.4
		Bagan Tiang	1	0.8	0.7	0.5	-6.7	4.5	-5.9
		Beriah	0.6	0.5	0.5	0.5	-10.7	10.5	36.5
		Gunong Semanggol	1.1	0.8	0.7	0.7	-12.9	1	21.3
		Kuala Kurau	2.2	1.6	1.3	1	-13.8	-5.3	1.6
		Parit Buntar	1.3	1.6	1.4	1.4	38.9	9.5	24.6
		Selinsing	0.5	0.4	0.4	0.4	-3.9	4.4	19.5
		Tanjong Piandang	1.1	0.7	0.6	0.5	-17.5	0.1	-6.1
	Larut dan Matang	*Asam Kumbang*	5	4.2	3.6	3.6	-0.8	3	26.1
		Bukit Gantang	0.9	0.7	0.6	0.5	0.6	-6.9	1.3
		Jebong	1	1	0.8	0.8	18	-11.3	23.5
		Pengkalan Aor	0.8	1.3	1.4	1.4	85.5	29.9	24.2
		Simpang	0.3	0.3	0.2	0.2	19.8	-17.2	19.4
		Sungai Limau	0.2	0.2	0.1	0.1	9.8	-7.9	13.6
		Terong	0.3	0.2	0.2	0.2	-10.6	-8.8	11.7
		Tupai	1.9	1.8	1.4	1.3	9.7	-3.3	14.1
Kuantan	*Kuantan*	*Beserah*	2.5	2.1	2	2.6	33.1	22.4	65.9
		Kuala Kuantan	**47.2**	**44.6**	**45.8**	**44.9**	**46.6**	**34.4**	**25.2**
		Penor	1.1	1	1	1	45.4	28.5	33.9
		Sungai Karang	3.3	4.5	5.4	7.3	108.4	57.7	73.4
	Pekan	*Ganchong*	0.4	0.3	0.2	0.2	-5.3	10.8	33.9
		Kuala Pahang	1.8	1.3	1.2	1.1	16.3	15.7	16.1
		Langgar	0.8	0.7	0.7	0.8	30.4	37.2	44.6
		Pahang Tua	0.9	1.3	1.9	1.6	118.4	85.4	7.1
		Pekan	6.5	5	4	3.6	18	6	15.4
		Pulau Manis	0.4	0.3	0.2	0.3	2.9	17.4	41.2
		Pulau Rusa	0.2	0.1	0.1	0.1	-3.7	-12.5	19.4
		Temai	0.3	0.2	0.2	0.1	-8.5	21.6	-8.5
	Dungun	*Kuala Abang*	0.4	0.6	0.6	0.6	171.3	33.2	15
		Kuala Dungun	9.5	7.2	5.8	4.6	18.9	4.8	0.1
		Kuala Paka	2.4	2.8	3.9	4.1	81.8	81.5	33.4

		Kumpal	0.6	0.5	0.5	0.4	30.1	18.8	2.8
		Rasau	0.8	3.7	3.3	1.7	631.2	17.3	-32.8
		Sura	2.4	3.1	3.5	5.2	99.1	50.8	88.2
	Kemaman	*Banggul*	0	0.2	0.9	0.9	912	471.3	32.5
		Binjai	1.1	1.4	1.4	1.9	92.1	35	69.1
		Cukai	8.7	8.1	7.3	7.3	44.2	18.1	27.6
		Hulu Cukai	0.6	1.5	1.5	1.2	250.6	37.9	-0.6
		Hulu Jabur	1.6	1.4	1.2	0.9	38.1	7.1	-6.6
		Kemasik	1.4	1.2	1.2	0.9	29.1	35.9	-3.6
		Kerteh	2	3.9	2.7	3.2	195.3	-8.1	50.9
		Kijal	1.7	1.7	1.5	1.4	50.3	13.3	19.5
		Pasir Semut	0.5	0.4	0.3	0.3	14.3	14.7	18.6
		Teluk Kalung	0.6	1.1	1.7	1.8	185.2	92.6	36.5
Johor Bahru	*Johor Bahru*	***Bandar Johor Bahru***	**30.9**	**20.4**	**12.8**	**7.7**	**11.6**	**-4.7**	**-12.9**
		Plentong	20.2	27.3	31.3	30.5	128.1	74.4	41.4
		Pulai/Jelutong	7.1	16.2	20.1	23.2	285.2	89	67
		Sungai Tiram	2.1	1.2	0.7	0.8	-6	-13.6	82.6
		Tanjung Kupang	1.1	0.8	0.7	0.7	25.7	42.6	31.5
		Tebrau	15.5	15.4	17	19.5	67.9	68.2	66.3
	Kulaijaya	*Sedenak/ Bukit Batu*	5.1	3.6	2.6	1.9	22.1	8.3	7.6
		Senai/Kulai	11.4	11.1	11.7	13.2	64.7	60.2	64.1
	Pontian	*Air Masin*	1.2	0.7	0.5	0.3	-2.1	1.6	-7.3
		Jeram Batu/ Pengkalan Raja	3.4	2.1	1.8	1.6	5.2	26.9	33.8
		Serkat	1.6	1	0.7	0.5	4.5	8.4	-0.8
		Sungai Karang	0.5	0.3	0.2	0.1	-5.4	3.5	-10.3

*Note: *The reason Putrajaya share and growth was 0 is that it was part of Sepang District before it was separated into a Federal Territory in 2001.*
Bold is mukim containing the largest city in each metropolitan region.

Kuala Lumpur Metropolitan Region –Tables 4.3 and 4.4 show that the dominance of *Kuala Lumpur* has deteriorated since 1980 due to significant population growth in the adjacent districts. The population share in adjacent districts—*Petaling* and *Hulu Langat*—significantly increased, from 18.5 percent to 28.1 percent and from 9.1 percent to 18.1 percent, respectively. In contrast, *Kuala Lumpur*'s share rapidly declined, from 47.2 percent to 25.3 percent. At the *mukim* level, *Bandar Petaling Jaya, Damansara*, and *Sungai Buloh* contributed to the high population share of *Petaling District* while *Hulu Langat District*'s share was due to the *mukim* of *Ampang* and *Kajang*. These *mukim* comprise many new townships that grew rapidly throughout the study period. According to Abdullah and Mohd (2009), the population in *Petaling* and *Hulu Langat* grew eight and seven times faster, respectively, than *Kuala Lumpur* during the period of 1991-2000. Surprisingly, the share of *Petaling District* exceeded *Kuala Lumpur*'s in 2010, indicating that this district has a higher population concentration than *Kuala Lumpur*.

As discussed in the previous section, *Selangor State* received an overflow of urban development from *Kuala Lumpur*, and these small areas (*mukim* particularly) received a huge concentration of the population. Many new townships were established and multinational companies, residential areas, commercial and industrial centres, and major infrastructure were built due to the overflow of urban development from *Kuala Lumpur*. Kuala Lumpur City Hall had raised concerns in *Kuala Lumpur Structure Plan 2020 Report (2003)* that the decline of the population is one of the biggest problems facing the region (Dewan Bandaraya Kuala Lumpur, 2003). *Kuala Lumpur's* mayor also expressed the view that the city is practically dead once office workers leave for their homes in suburban areas (Shuid, 2004). Figure 4.3 shows the overall growth of the Kuala Lumpur Metropolitan Region from 1980 until 2010.

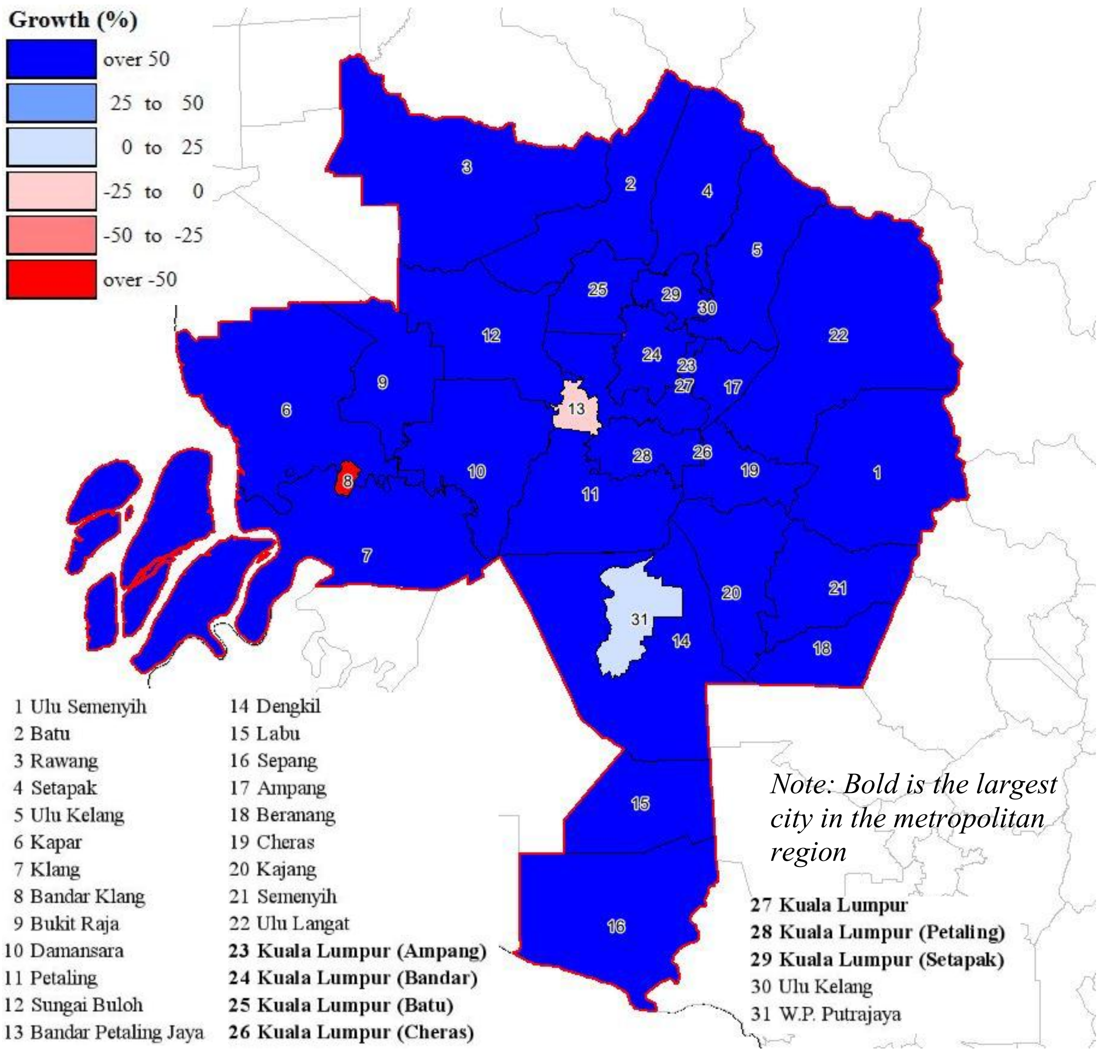
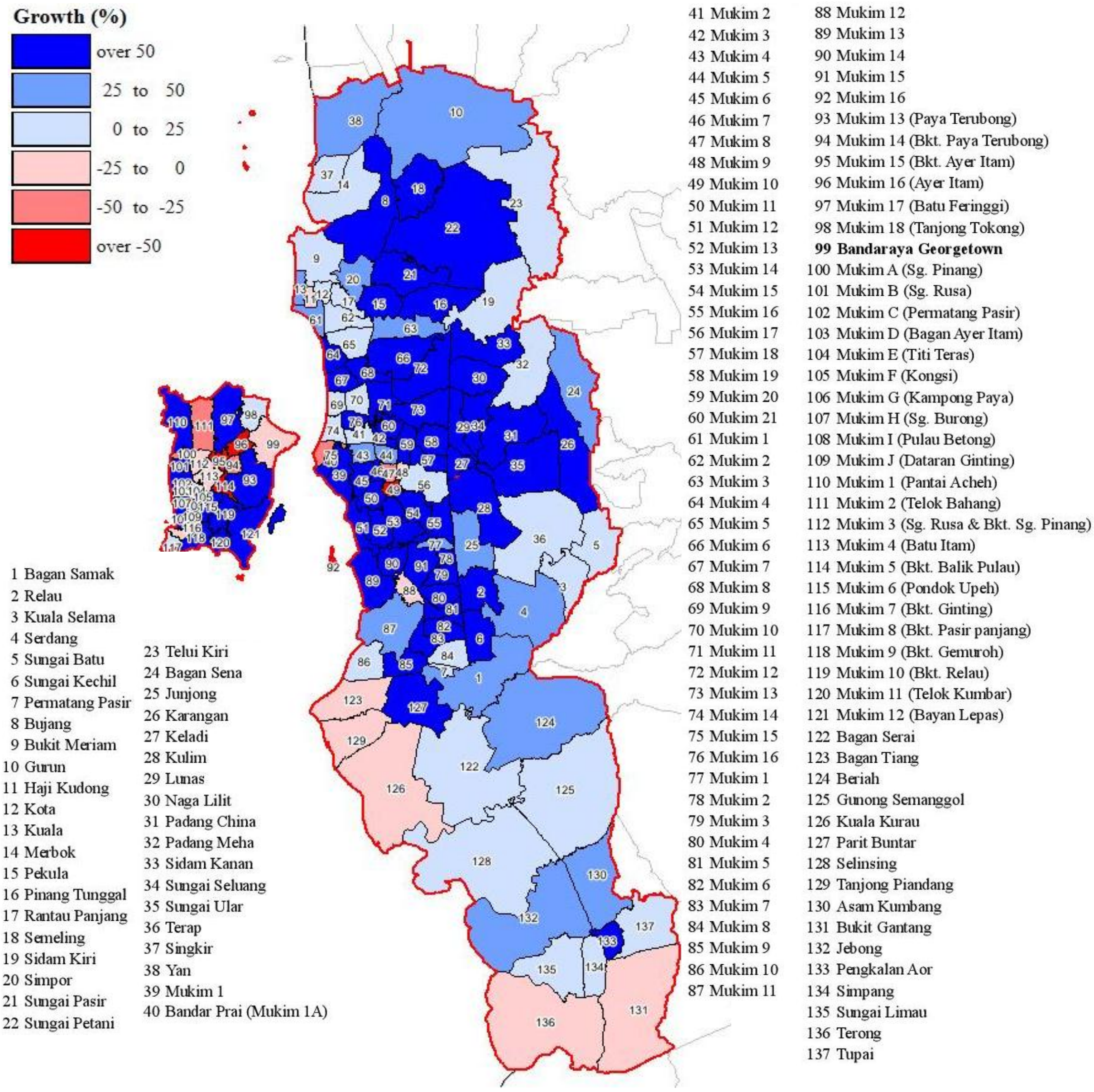


Figure 4. 3: Population growth in Kuala Lumpur Metropolitan Region, 1980-2010

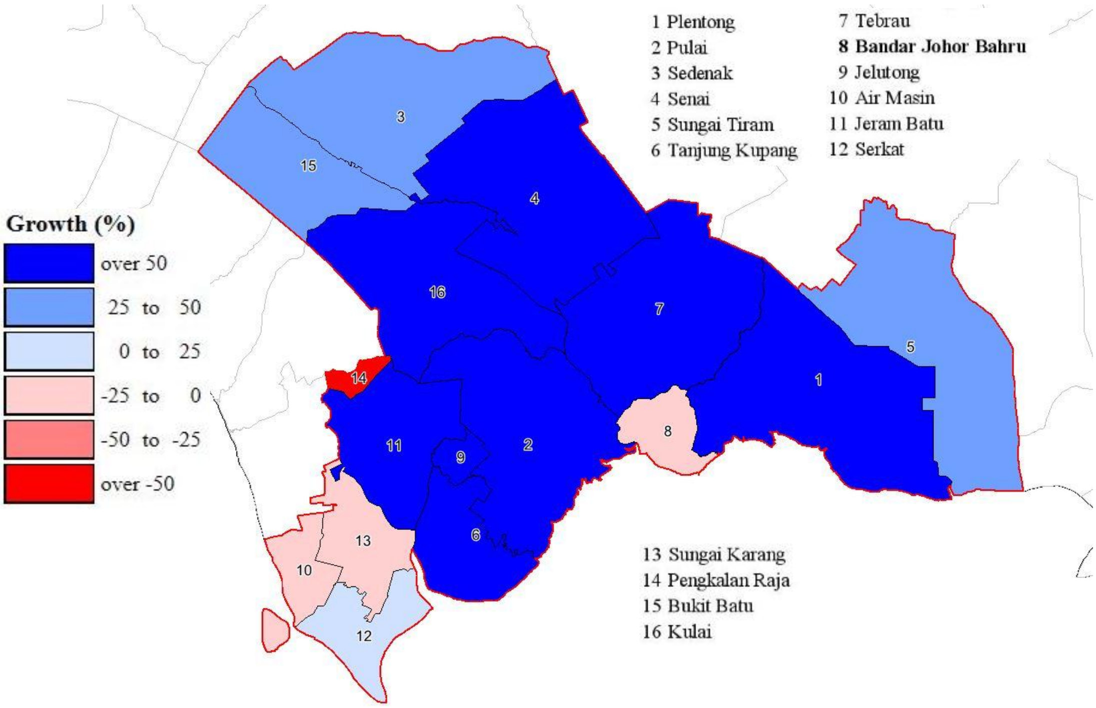
Georgetown Metropolitan Region –*Timur Laut and Kuala Muda* Districts saw significant change within the Georgetown Metropolitan Region*.* *Timur Laut* includes the second largest city in Malaysia, *Bandaraya Georgetown Mukim*. However, the population share in this city declined significantly, from 15.4 percent to 7.3 percent. In contrast, the share of a neighbouring area*, Mukim 13* (also known as *Paya Terubong*), increased from 4.7 percent to 8.6 percent, thus exceeding *Bandaraya Georgetown*'s share in 2010. Similar to *Kuala Lumpur*'s experience (described in the previous subsection), rapid suburbanisation and decentralisation of urban development in adjacent areas caused urban expansion away from the city centre. Furthermore, the establishment of manufacturing townships in these areas in 1980 fuelled the deconcentration process (Abdullah et al., 2009). Geographically, this metropolitan region covers the whole island of *Pulau Pinang State* (which includes the second largest city in Malaysia) and several areas in mainland Peninsular Malaysia. These areas are connected by a bridge and main highway, allowing people to commute to the city. To date, the suburbanisation process has expanded towards the mainland rather than on the island due to cheaper land prices (Abdullah et al., 2009). Figure 4.4 shows the overall growth of Georgetown Metropolitan Region from 1980 until 2010.



Note: Bold is the largest city in the metropolitan region

Figure 4. 4: Population growth in Georgetown Metropolitan Region, 1980-2010

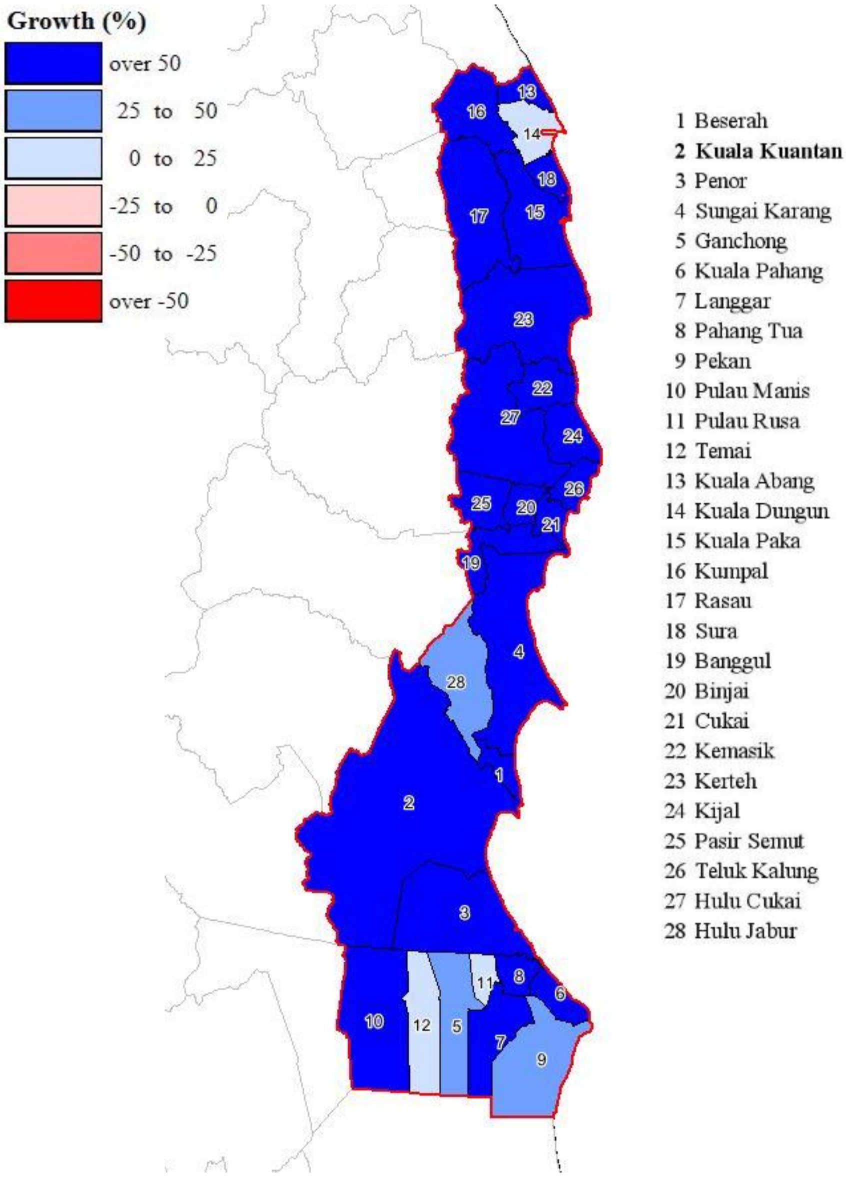
Johor Bahru Metropolitan Region – Unlike the previously mentioned metropolitan regions, *Johor Bahru* District dominates the population share and growth in this region. This district includes the third largest city, *Bandar Johor Bahru Mukim*. The largest city, however, experienced a major decline in the population share, from 30.9 percent to 7.7 percent, over the three decades examined. On the contrary, the population share of surrounding areas (*Plentong, Pulai/Jelutong, Tebrau,* and *Senai/Kulai)* were two to four times larger than that of the largest city in 2010. This is due to the relocation of the state administrative centre from the city centre to the adjacent area of *Pulai/Jelutong* and the establishment of main industrial areas and ports in *Plentong*. The increase in population growth in *Tebrau* and *Senai/Kulai*, was due to large-scale urban development, including a university (*Universiti Teknologi Malaysia*) and new townships. Figure 4.5 shows the overall growth of Johor Bahru Metropolitan region from 1980 until 2010.



Note: Bold is the largest city in the metropolitan region

Figure 4. 5: Population growth in Johor Bahru Metropolitan Region, 1980-2010

Kuantan Metropolitan Region – Similar to Johor Bahru Metropolitan Region's experience, *Kuantan* District dominated the population share and growth in Kuantan Metropolitan Region. This was mainly due to *Kuala Kuantan Mukim*, which includes the fourth largest city, *Kuantan*. Unlike the previously discussed regions, the largest city in this region maintained a strong population share of almost 50 percent, and there was no sign of population deconcentration into neighbouring *mukim*. Other *mukim* that saw significant growth are *Sungai Karang, Pekan, Kuala Dungun, Sura*, and *Cukai*. These *mukim* consist of self-sustaining small cities located far from the largest city. Figure 4.6 shows the overall growth of Kuantan Metropolitan Region from 1980 until 2010.



Note: Bold is the largest city in the metropolitan region

Figure 4. 6: Population growth in Kuantan Metropolitan Region, 1980-2010

What lessons can these detailed patterns of metropolitan change teach us regarding the course of urbanisation in Malaysia? In summary, the *mukim* that include the largest cities (*Kuala Lumpur, Bandaraya Georgetown,* and *Bandar Johor Bahru*) appear to have experienced slow and or negative population growth. In contrast, the neighbouring *mukim* exceeded the population share and growth of these large cities. Deconcentration of urban development and urban sprawl within metropolitan regions are the major causes of this situation (Abdullah et al., 2009). As these cities grew rapidly, urban development overflow into adjacent areas resulted in the establishment of many new towns and industrial centres away from large cities.

Other factors include the relocation of national and state administration centres within metropolitan regions to avoid urban congestion in city centres and establishment of large-scale development in previously undeveloped areas to encourage urban development and growth away from urban centres (Hasan & Nair, 2014).

Furthermore, observation at the district level may not be adequate to examine population change in detail. For example, previous studies (see Abdullah, 2012; Abdullah & Mohd, 2009; Osman, Nawawi, & Abdullah, 2009) identified *Timur Laut District* as one of the largest cities in Malaysia, which is not entirely true because the largest city (*Bandaraya Georgetown Mukim)* covers only a small portion of the district while the rest of the district is rural in nature.

While it is clear that metropolitan areas are spreading, reflecting ongoing urbanisation in Malaysia, the relevance of these trends to differential urbanisation theory depends on what is happening in other settlements and rural areas and on migration patterns. A major shortcoming of this section is that population change is observed only in metropolitan regions. As discussed in Chapter 2, urbanisation is caused by the deconcentration of the population in lower-level settlements and rural areas. Therefore, the next section tackles this gap by examining population change for all settlement types through the application of differential urbanisation theory.

4.4 Population growth by all settlement types

This section tests differential urbanisation theory by analysing population change using the new urban-rural classification (built using *mukim* units) developed in Chapter 3 (Figure 3.6). Figure 4.3 shows population growth and shares by all settlement types in Malaysia from 1980-2010.

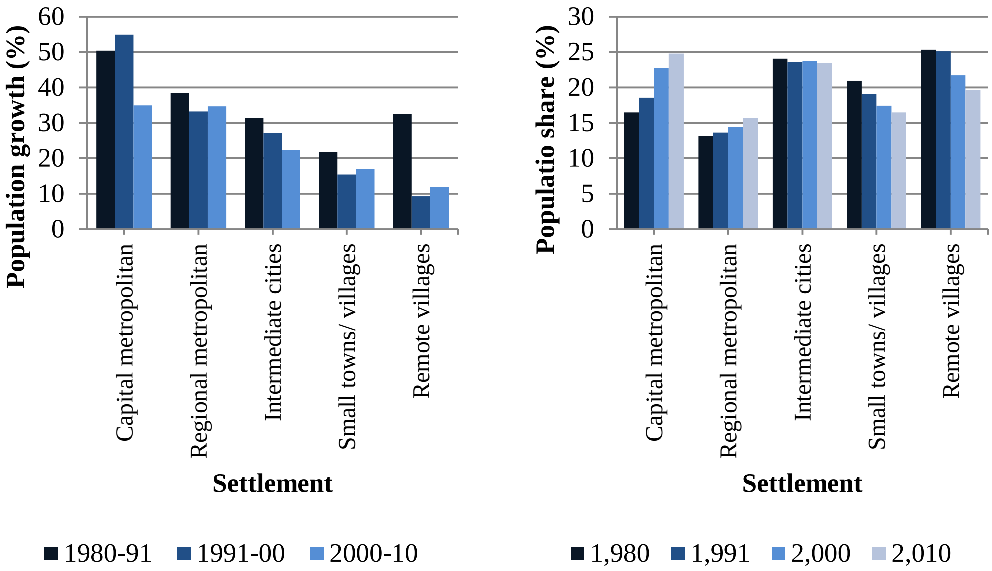


Figure 4. 7: Population growth and shares by settlement type, 1980-2010

Based on Figure 4.7, the capital metropolitan settlement type appears to have seen the largest population growth, followed by settlements lower in the hierarchy. In other words, growth occurs according to hierarchical order; the larger the settlement, the larger the expected growth. However, the results for population share tell a different story. The population share increased in larger cities (capital and regional metropolitan) but decreased in lower-level settlements (small towns/villages and remote villages).

Another way to observe population change is by examining the relative change rate to the starting year population and overall change in Malaysia (Figure 4.4).

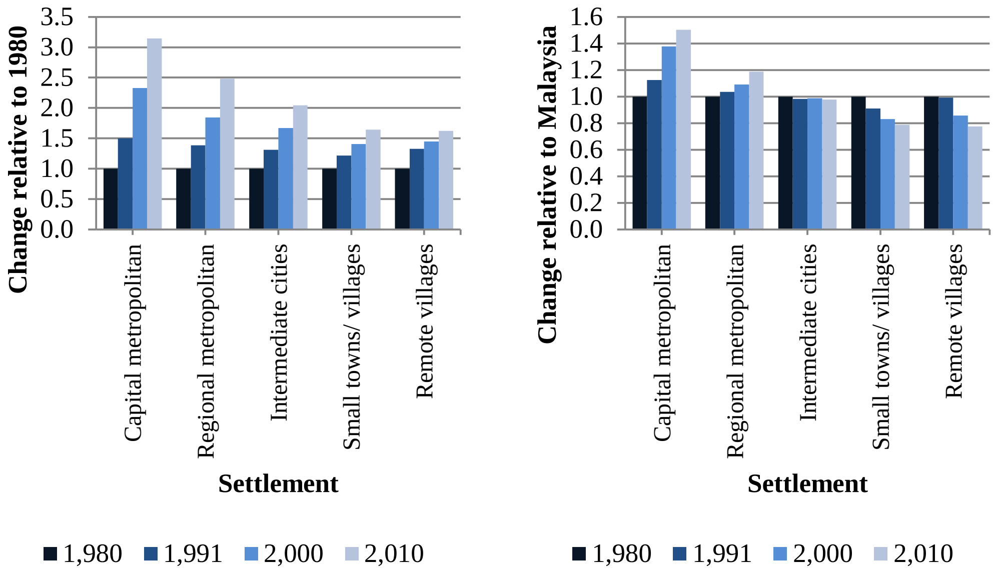


Figure 4. 8: Change relative to 1980 population and to Malaysia as a whole

Similar to the results for population growth, the relative change rate in Figure 4.8 displays a similar pattern; the larger the settlement, the larger and faster the expected change in population. Comparison to overall change in Malaysia, only the metropolitan cities (capital and regional) displayed a positive increase in population, while the population of other settlement types declined rapidly, except intermediate-sized cities. This situation indicates that overall change in Malaysia since 1980 was primarily due to population growth in metropolitan cities, primarily in capital metropolitan cities.

Metropolitan cities are a combination of a core city and its surrounding suburban areas, so the next step is to disaggregate them. Figure 4.5 shows population growth and shares of core cities and suburban areas for metropolitan cities.

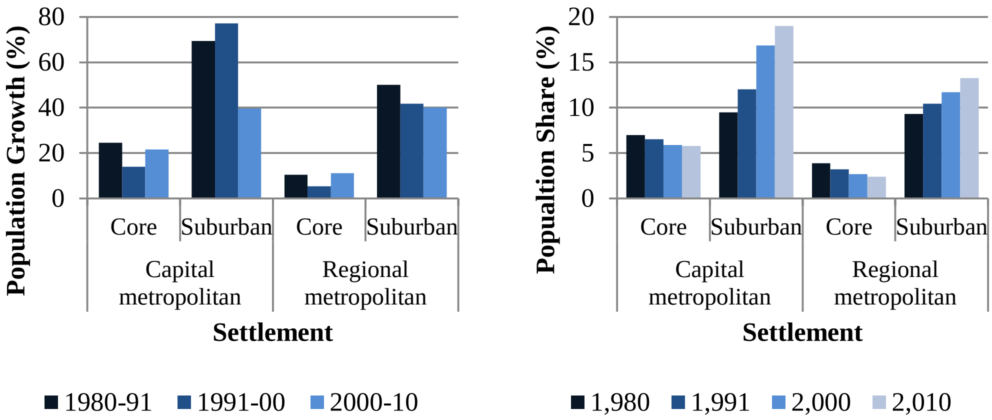


Figure 4. 9: Population growth and share in metropolitan cities.

Figure 4.9 clearly shows that the population growth in suburban areas of both capital and regional metropolitan cities is significantly higher than in the core cities, despite decreasing over time. In terms of population share, the core and suburban areas in both types of metropolitan area display a contrasting pattern; the share in suburban areas significantly increased while the share in core cities decreased. The results show that suburban areas gained the largest population concentration, mainly due to rapid suburbanisation since 1980. Furthermore, similar results emerged in terms of relative change to the starting year population and to overall change in Malaysia, as shown in Figure 4.6.

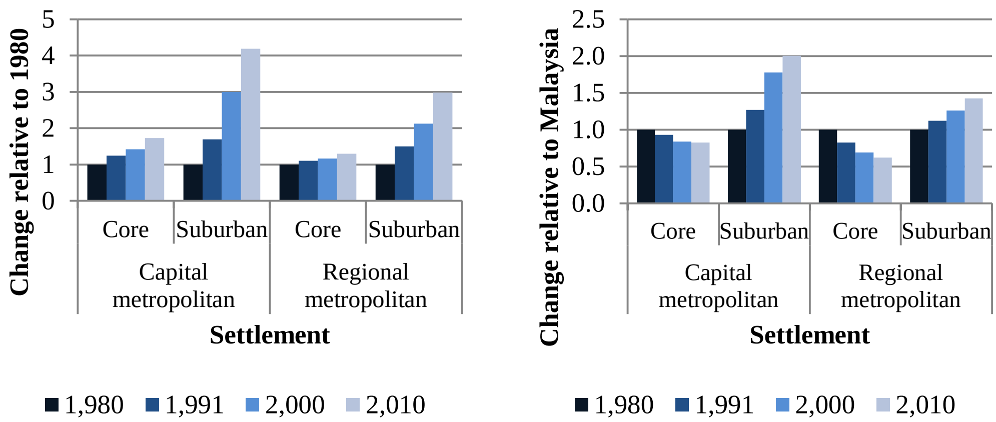


Figure 4. 10: Change relative to 1980 population and to Malaysia as a whole in metropolitan cities

Figure 4.6 shows the population in suburban areas was four times larger in 2010 compared to the starting year population in 1980. Further, suburban areas were the sole contributors to overall population change in Malaysia. These results indicate the rapid suburbanisation process was a major factor in the overwhelming population growth in metropolitan cities.

In relation to differential urbanisation theory, by 1980, Malaysia may have reached the final stage of urbanisation, APC, due to high successive growth in suburban areas and the decline of the primacy of primate cities. However, it is too early to determine the urbanisation stage because the theory mainly concerns migration flows. Population growth between settlements hence requires further analysis (in Chapter 5). According to UN-Habitat (2016), there have been major patterns of suburbanisation in the developing world for the past few decades. These patterns have been observed in many developing countries, and the trend is motivated by individual preferences (e.g., lower living costs, proximity to workplaces), poor land management, and the increase in mobility caused by the development of highways and the dominance of private transportation. This is true in the Malaysian context, where recent studies have shown urban sprawl has become a major problem due to rapid suburbanisation in metropolitan regions caused by uncontrolled urban development and growth (Abdullah, 2012; Hasan & Nair, 2014).

To expand the current results, the next section investigates the demographic and socio-economic changes for each settlement type. The aim is to shed some light on the nature and potential drivers of population change that have been identified. Further, because differential urbanisation theory has never been tested for Malaysia, the next section also represents an original contribution of this thesis.

4.5 Demographic and socio-economic changes by settlement type

Due to issues regarding the availability of data mentioned in Section 4.2, for most of the analyses in this section, data was taken from IPUMS International rather than the Malaysia Department of Statistics. However, the data from the IPUMS International includes only 2 percent of the total district population and is only available from 1980 to 2000. Unlike the previous section, this section applied the settlement classification determined by district units in Chapter 3 to analyse demographic and socio-economic changes by settlement type. Change by gender and age group is discussed in Section 4.5.1, change by ethnicity is discussed in Section 4.5.2, change by education attainment is discussed in Section 4.5.3, change by occupation industry is discussed in Section 4.5.4, and change by type of occupation is discussed in Section 4.5.5.

4.5.1 Gender and age group

The results for gender and age group are split into two parts. The first part focuses on results based on a census from IPUMS International, and the second part focuses on results from census data provided by the Malaysia Department of Statistics. The results are separated into two parts because IPUMS International offers gender and age information by district level but cover only the period from 1980 to 2000. In contrast, the Malaysia Department of Statistics offers more recent census data that covers the period from 1980 to 2010 by *mukim* level but is limited by age group. This section discusses outcomes that show the importance of gender and age group by settlement type and assesses the similarities and differences of the results from the two data sources.

1. IPUMS International

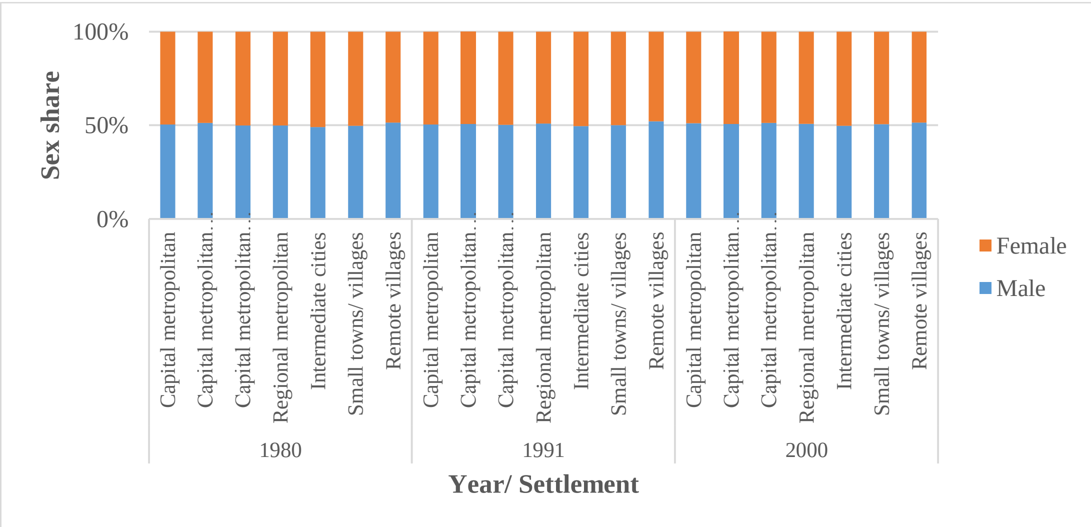


Figure 4. 11: Sex share by settlement type, 1980-2000

As shown in Figure 4.11, there is no major difference of sex by each settlement type. However, in terms of age group, Figures 4.12 and 4.13 show a noticeable pattern of population growth and shares.

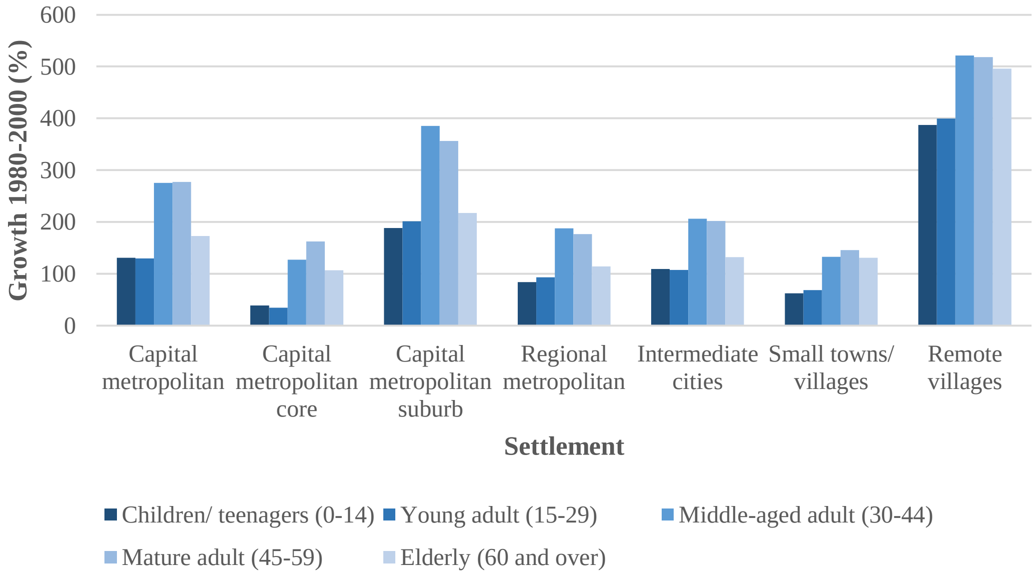
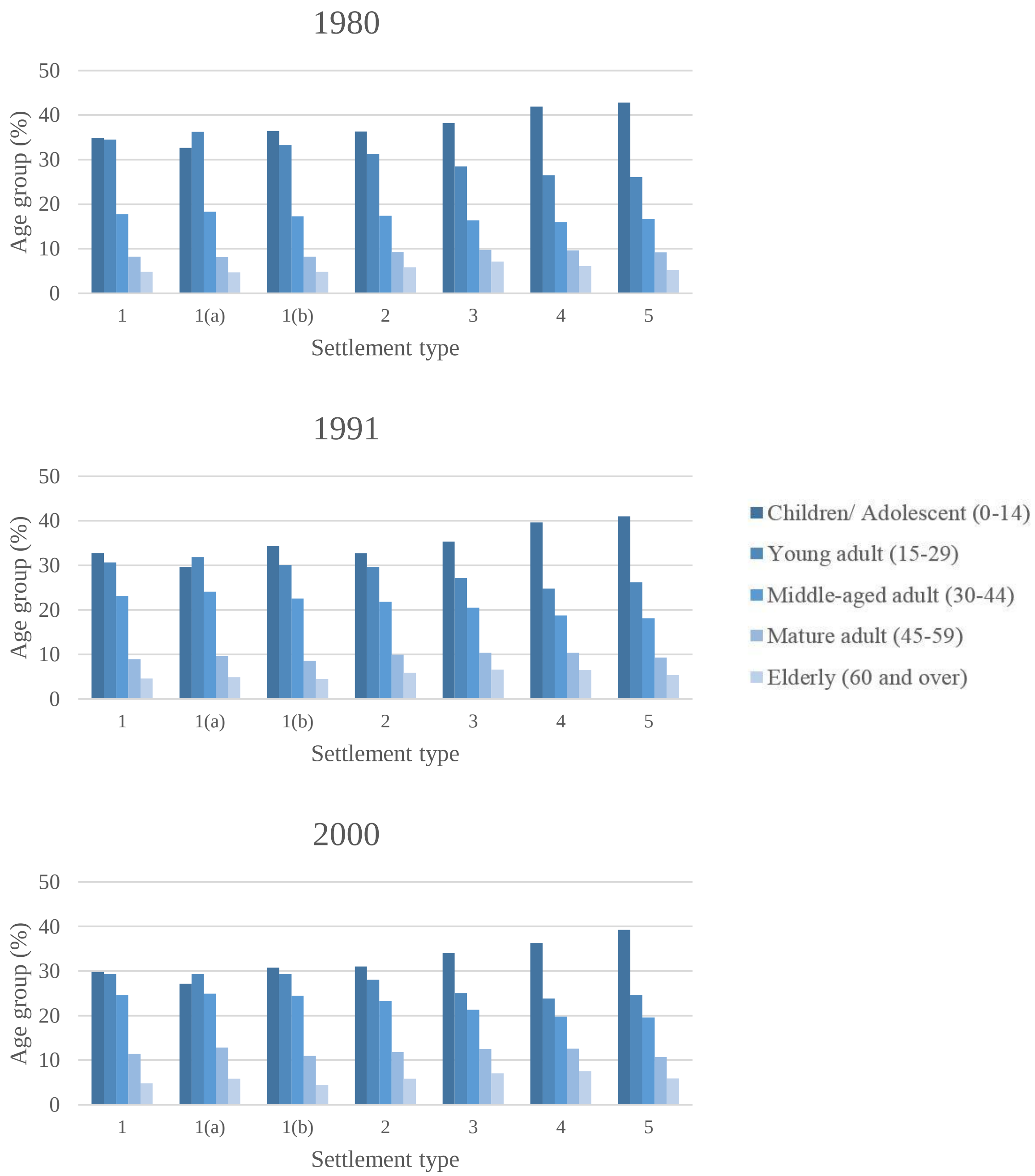


Figure 4. 12: Population growth by age group and settlement type, 1980-2000

Surprisingly, the results in Figure 4.12 indicate that remote villages had the highest growth for all age groups compared to other settlement types. This is followed by the capital metropolitan areas, intermediate-sized cities, regional metropolitan areas, and small towns/villages. As expected, the high growth in capital metropolitan areas was mainly due to the included suburban areas. The results also show that there was more growth in the older adult (30 years and over) population than for younger age groups (below 30 years) for all settlement types. To investigate this matter further, Figure 4.13 shows the change in population shares.



Note: 1 is Capital metropolitan, 1(a) is Capital metropolitan core, 1(b) is Capital metropolitan suburban, 2 is Regional metropolitan, 3 is Intermediate cities, 4 is small towns/villages, and 5 is Remote villages

Figure 4. 13: Population shares by age group and settlement type, 1980-2000

The population share shown in Figure 4.13 tells a different story compared to the population growth shown in Figure 4.12. First, the share pattern is similar over the years for most settlement types, with the younger age groups having larger shares than the older age groups. However, the shares for younger age groups gradually diminished in subsequent years. In contrast, the shares for older adults increased. These results indicate that the older age groups saw more significant growth than the younger age groups. However, although remote villages had the highest growth, there were no significant changes in their share.

2. Department of Statistics Malaysia

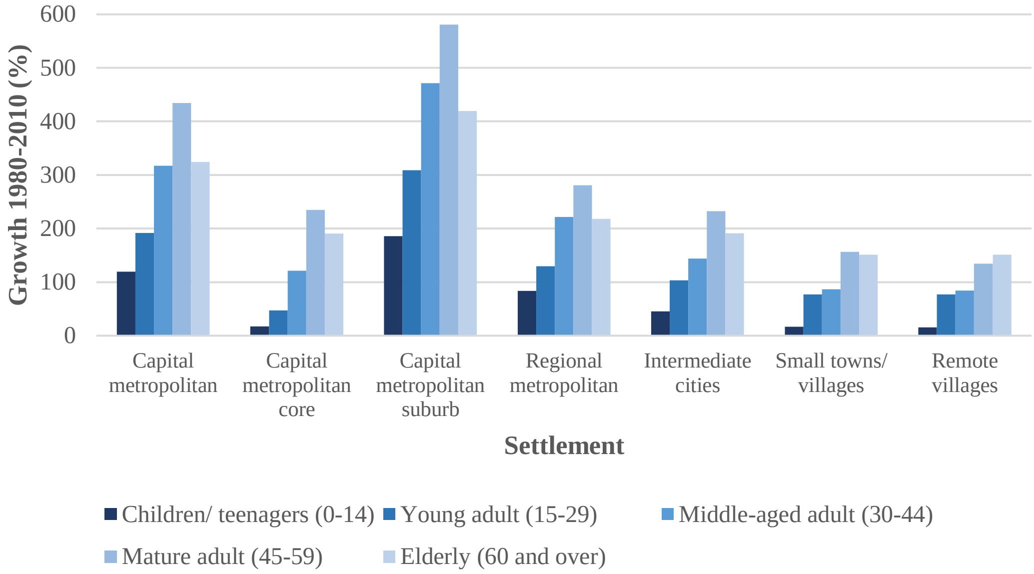
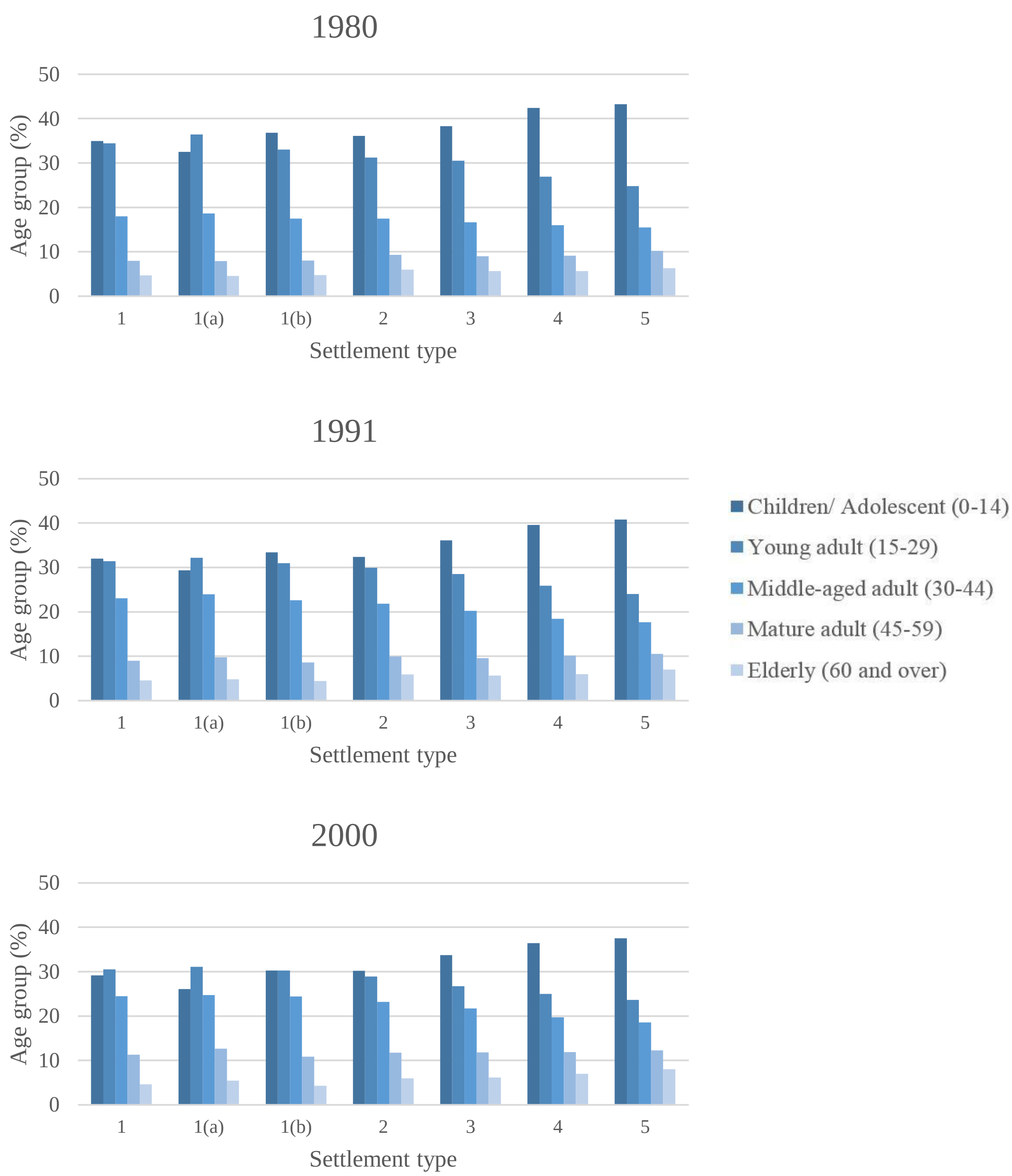
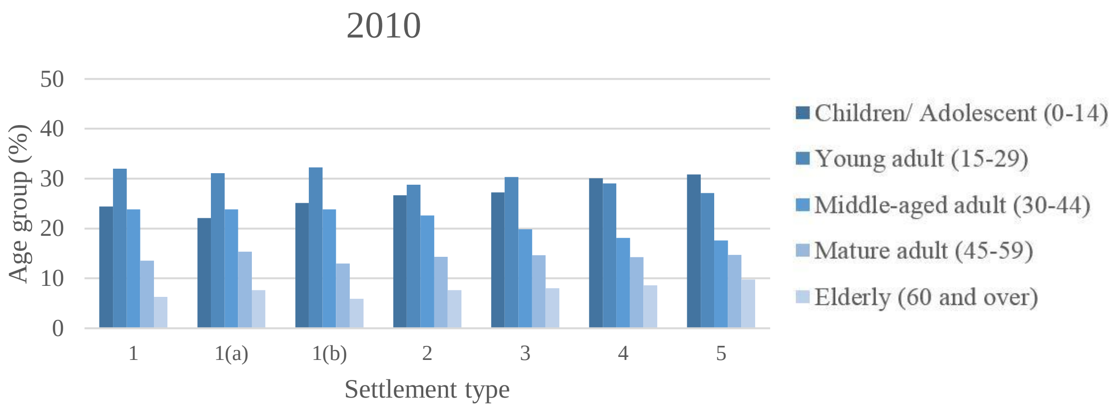


Figure 4. 14: Population growth by age group and settlement type, 1980-2010

Contrary to the growth results from IPUMS International data in Figure 4.12, the growth results from Malaysia Department of Statistics data in Figure 4.14 display a more sensible pattern – the larger the settlement type, the higher the growth. One of the main reasons for this is the availability of data between these two sources. The data from the Department of Statistics Malaysia covers the overall Malaysia population, and the analysis was conducted using *mukim* units to analyse change between settlement types, which is more detailed and accurate than the previous analysis that is done by district level (Figure 4.12). Still, growth for each age group displays a fairly similar pattern; the older the age group, the higher the growth. Note that the growth percentages are also different between Figure 4.14 and Figure 4.12. This difference between results is due to Modifiable Area Unit Problem (MAUP). MAUP is a situation when the boundaries of geographical units can be changed which could lead to inconsistent results in statistical analysis (Wong, 2004). It appears that there is an aggregation effect because both boundaries are at different scale and is used throughout the period studied, 1980-2010.





Note: 1 is Capital metropolitan, 1(a) is Capital metropolitan core, 1(b) is Capital metropolitan suburban, 2 is Regional metropolitan, 3 is Intermediate cities, 4 is small towns/villages, and 5 is Remote villages

Figure 4. 15: Population shares by age group and settlement type, 1980-2010

Despite the differences in data, the results for shares in Figure 4.15 agree with the shares in Figure 4.13. First, the age group distribution for all settlement types resembles a pyramid, where the older the age group, the smaller the share. Second, there is a diminishing share for younger age groups and an increasing share for older adults.

Overall, there are four main findings in this section. First, change by age group is more important than change by sex. Second, despite the differences between IPUMS International data and Malaysia Department of Statistics data and different spatial scales (district/*mukim)* used to examine settlement change, the results agree. In other words, district-level data from IPUMS International produces fairly similar results to the *mukim*-level data from the Malaysia Department of Statistics for explaining settlement change in Malaysia. Third, the increasing growth and shares for older age groups indicate that more people in these age groups are surviving. One of the main reasons for this the continuous increase of life expectancy and decline in mortality levels since the 1950s due to the improvement of nutrition, the availability of preventive health programs, and greater accessibility to curative medicine (Hirschman, 1980). In contrast, the diminishing growth and share of younger age groups are due to the continuous decline in fertility levels since the 1960s caused by social change, improved education, women's empowerment and increasing participation in the working sector, and the postponement of marriage and childbearing (Hirschman, 1980). Fourth, the decline of fertility and mortality, followed by the rise of population growth and shares for older adults may relate to population momentum. Population momentum arises when previous high fertility results in a large proportion of the female population being of reproductive age, hence leading to a high birth rate (Keyfitz, 1971).

4.5.2 Ethnicity

Besides gender and age, ethnicity also plays an important role in population distribution in Malaysia. The population in Malaysia is comprised of three major ethnicities: *Bumiputera*, Chinese, and Indian. *Bumiputera* is the largest ethnic group and consists of *Malay*, who are the majority and indigenous people.

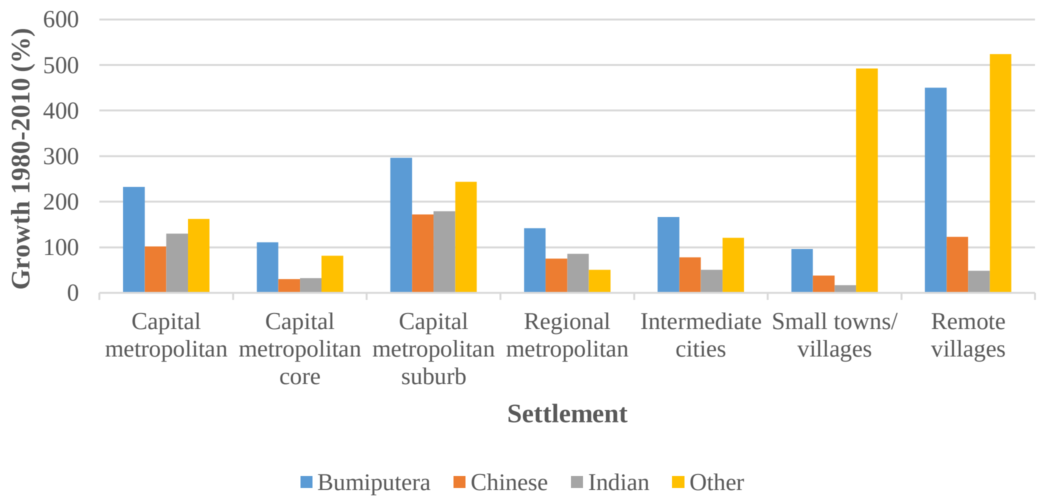


Figure 4. 16: Population growth by ethnicity and settlement type, 1980-2000

Figure 4.16 clearly shows that the *Bumiputera* had the highest growth, especially in larger cities. In contrast, other ethnic minorities had the highest growth in small towns/villages and remote villages. There are two possible reasons for this. First, due to racial tensions between ethnic groups, in 1969 the Malaysian government introduced the first national economic policies to restructure the community and eradicate poverty among these groups in the late 1960s by encouraging them (especially the *Bumiputera*) to become involved in commercial and trading activities in cities (Yaakob et al., 2010). Second, the high population growth for other ethnic minorities may have been due to the large numbers of ethnic minorities living in East Malaysia (Sabah and Sarawak States). According to Mahari (2011), more than 70 ethnic groups have been identified in Malaysia, including the three major ethnic groups.

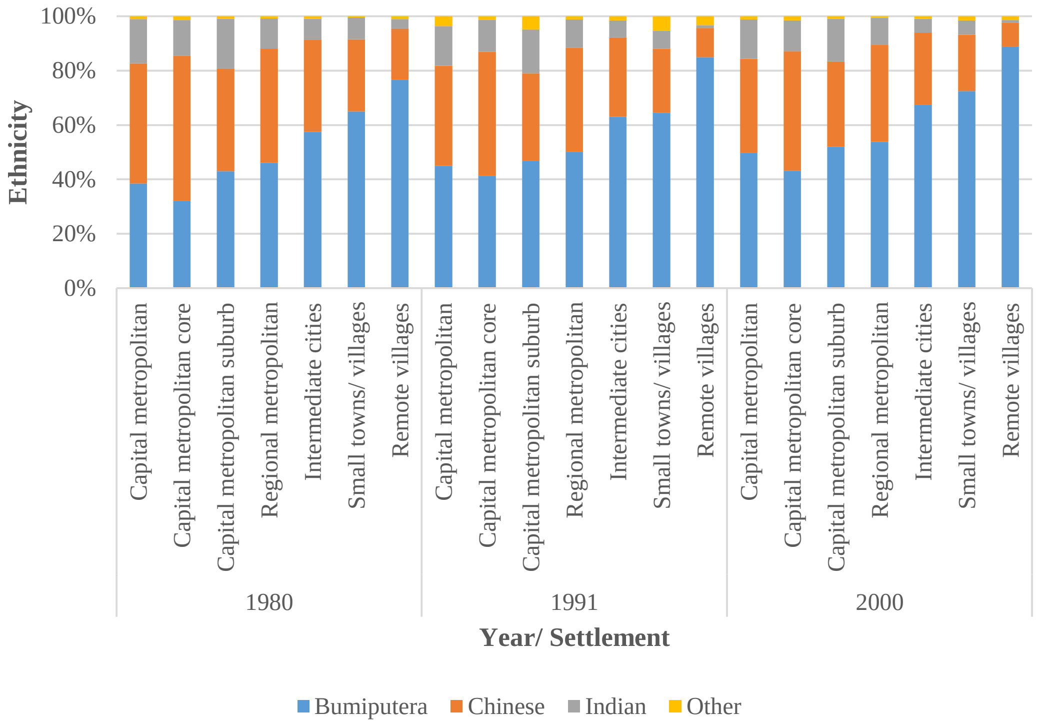


Figure 4. 17: Population shares by ethnicity and settlement type, 1980-2000

In terms of the population share, Figure 4.17 shows a clear distribution pattern between the *Bumiputera* and Chinese. The *Bumiputera* are more prominent in smaller settlements and rural areas. In contrast, the Chinese are more dominant in larger settlements. This is an end result from the 'divide and rule' strategy imposed by British colonizers before the country gained independence implemented to strengthen their cause and exploit the country by segregating ethnic groups in different places. Historically, there has been significant immigration from China and India due to the increasing demand for war-related workers during the Second World War (Lestari, 1997). The *Bumiputera* were redistributed by the colonizer to rural areas to focus on farming and agricultural activities while the minority Chinese and Indian focused on trade and business activities in towns and areas nearby areas.

Furthermore, the exceptional percentage of ethnic minorities in cities resulted from a colonial housing relocation programme meant to block communist involvement in local communities (Yaakob et al., 2010). Although the *Bumiputera* had higher growth than other ethnic minorities in cities (Figure 4.16), their share was still low, indicating that segregation still existed in 2010. However, on a positive side, the national economic policies mentioned earlier seem to have had positive impacts, as evidenced by high population growth and increasing shares of the *Bumiputera* in cities throughout the study period.

4.5.3 Education attainment

Education level is another important indicator for examining the population distribution between settlement types. Generally, those who have attained higher education tend concentrating in areas with higher economic activity. Figures 4.18 and 4.19 show population growth and shares in terms of education attainment by settlement type for 1980-2000, respectively.

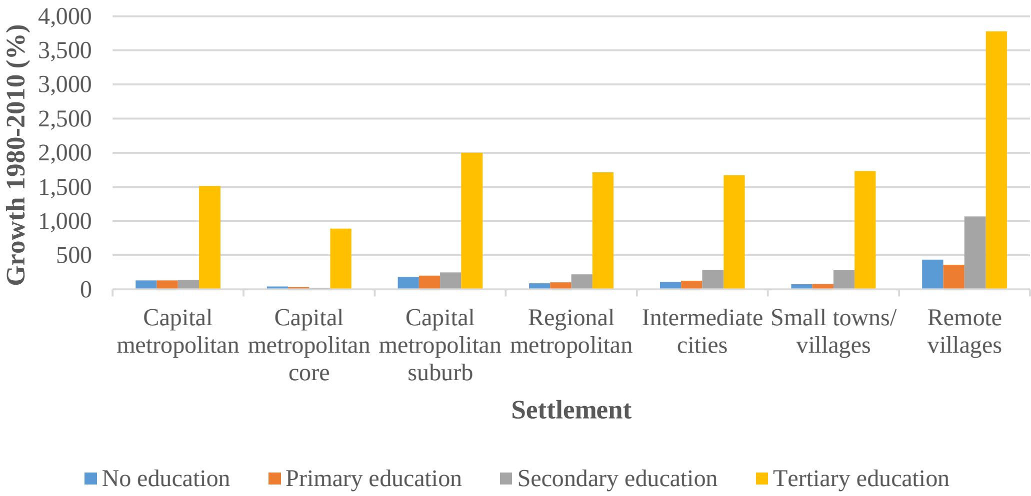


Figure 4. 18: Population growth by education attainment and settlement type, 1980-2000

Figure 4.18 presents two findings. First, surprisingly, remote villages had the highest population growth for all levels of education attainment. Second, the growth patterns do not vary by settlement type; the highest growth was seem among those who attained tertiary education for all settlement types. To further examine this, Figure 4.19 shows population shares by education attainment and settlement type.

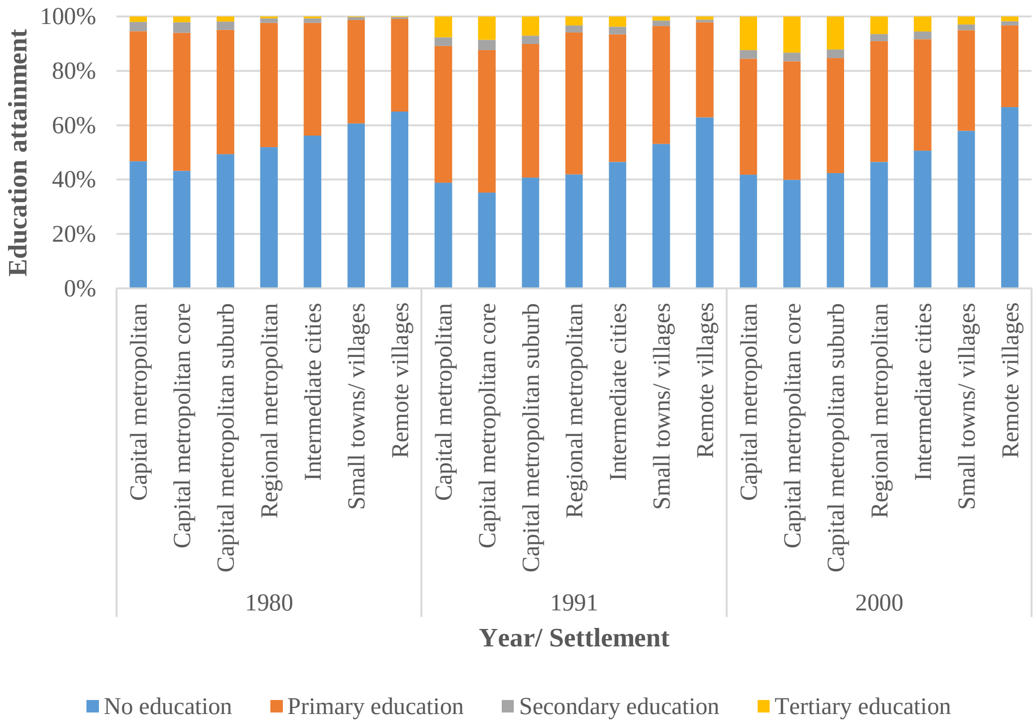


Figure 4. 19: Population shares by education attainment and settlement type, 1980-2000

Figure 4.19 exhibits a classic pattern; the larger the city, the more educated the population. The exceptional increase in population shares among those who attained tertiary education likely occurred because more universities and colleges were built during the study periods and hence more students were enrolled. However, although remote villages had the highest growth, there was no significant change in their population share over the period.

4.5.4 Occupation industry

The types of industry are commonly explained according to three main economic activities: primary, secondary, and tertiary. Primary industries involve extracting raw materials and include agricultural, fishing, forestry, and mining activities. Secondary industries consist of manufacturing and construction activities. Finally, tertiary industries include services such as electricity, gas and water, wholesale and retail trade, hotels and restaurants, transportation, storage and communications, financial services and insurance, public administration and defence, real estate and business services, education, health and social work, other services, and private household services.

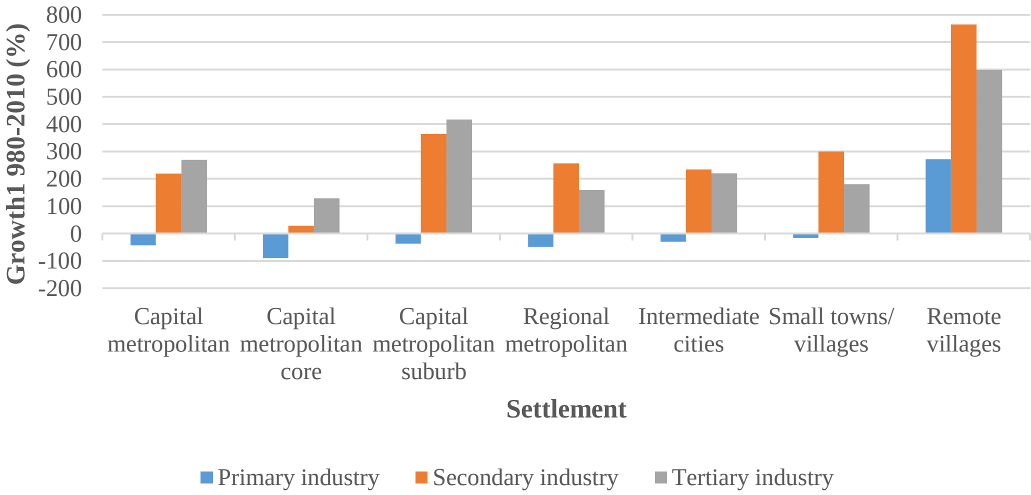


Figure 4. 20: Population growth by occupation industry and settlement type, 1980-2000

Figure 4.20 presents three findings. First, secondary and tertiary industries saw major growth compared to primary industry. The main reason for this is the economic transition from primary to multi-sector sector industries in the early 1980s. Manufacturing sectors and modern services grew substantially and were centralized in the vicinity of cities (Abdullah, 2003). Second, primary industry saw positive growth only in remote villages and not in other settlement types. This is unsurprising because these areas are dominated by agricultural activities since they are rural in nature. Third, similar as the previous sections, remote villages had the highest growth for all types of industry. However, this may not have affected the population share results and thus further results are provided in Figure 4.21.

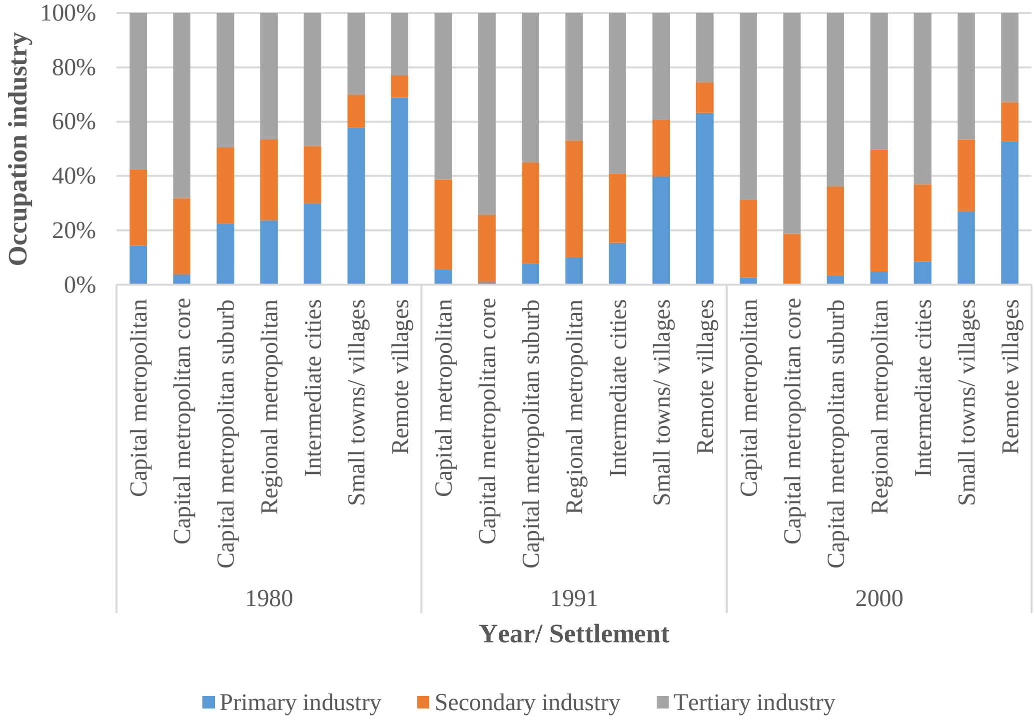


Figure 4. 21: Population shares by occupation industry and settlement type, 1980-2000

Based on the explanation in the previous paragraph, Figure 4.21 shows expected results for population shares — more tertiary and secondary industry economic activity in a larger settlement, and more primary industry activity in smaller settlements. Overall, this section shows a clear segregation of industries according to settlement type. To expand on this, the next section observes types of occupation.

4.5.5 Types of occupation

Generally, different areas offer different jobs, mainly is based on the economic potential of a particular area. For example, farming is dominant in rural areas while service-oriented jobs are commonly located in towns. To examine types of occupation in Malaysia, occupations are grouped into three main categories: high-paid jobs, middle-paid jobs, and low-paid jobs. High-paid jobs consist of legislators, senior officials, managers, and professionals. Middle-paid jobs consist of technicians, associate professionals, clerks, service workers, and shop workers and market sellers. Finally, low-paid jobs consist of skilled agricultural and fishery workers, crafts and related trades workers, plant and machine operators and assemblers, and elementary occupations.

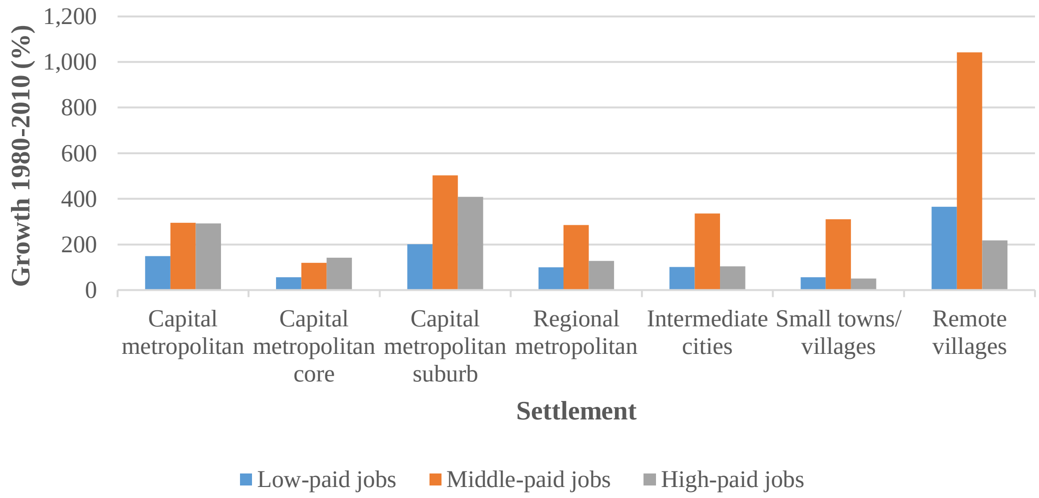


Figure 4. 22: Population growth by occupation type and settlement type, 1980-2000

Figure 4.22 displays a noticeable pattern. The population with middle-paid jobs had the highest growth in most settlement types over the study period, and the population share for those with high-paid jobs was higher in the capital metropolitan area (particularly in the core). This is expected considering the middle-paid jobs are commonly served anywhere in Malaysia, and high-paid jobs are commonly offered in larger cities. The capital metropolitan area is where many multinational companies are situated, with many firms and industries concentrated in this area.

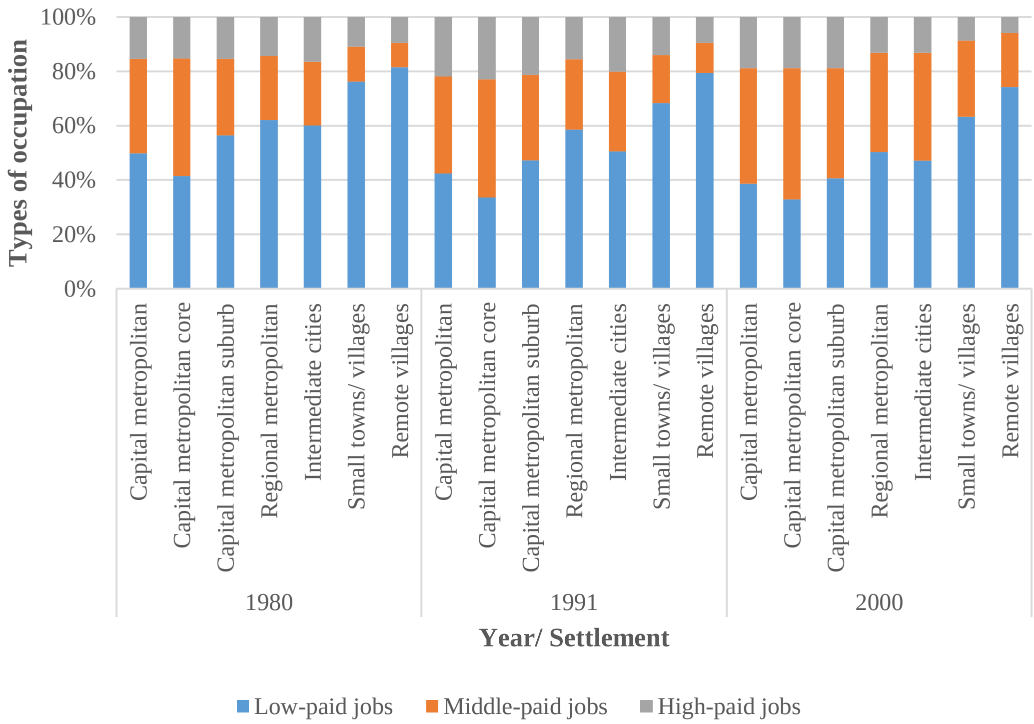


Figure 4. 23: Population shares by occupation type and settlement type, 1980-2000

Based on a similar interpretation as before, the population shares in Figure 4.23 show that middle-paid workers are more prevalent in cities while low-paid workers are dominant in smaller settlement types and rural areas for the whole study period. For high-paid workers, the share pattern is the same as for middle-paid worker; they are more concentrated in cities than in rural areas.

Overall, the results in this section are similar to those in previous sections. Occupation types and industry are associated with each other. For example, high and middle-paid workers normally provide services for the community, which is also known as tertiary industry. Low-paid or skilled workers are more dominant in primary industry such as agricultural and farming activities in rural areas.

4.6 Conclusion

Overall, this chapter offers three new contributions. The chapter extends the analysis of population growth and urbanisation in Malaysia beyond major urban areas to consider other areas (e.g., intermediate-sized cities, small towns/villages, and remote villages). Further, it provides an analysis of urban and rural areas using a classification relevant to differential urbanisation theory. Finally, it disaggregates areas by demographic and socio-economic factors for further analysis.

There are five important findings identified in this chapter. First, previous urbanisation studies in Malaysia have mostly been focused on certain small-area units, which are insufficient to examine the small-area changes. For example, existing studies (see Abdullah, 2012; Abdullah & Mohd, 2009; Osman, Nawawi, & Abdullah, 2009) identified *Timur Laut District* as one of the largest city in Malaysia, ignoring the fact that cities cover only a small portion of the district while the rest of it is rural in nature. Therefore, it is important, where possible, to observe population change at a smaller scale than the district level, such as by *mukim.* This is the first time this has been done for Malaysia.

Second, at the macro level, the large population concentration in the Central Region has created problems such as imbalanced urban development and growth since before the 1980s. The distribution of urban centres was spatially uneven, and most cities were located in higher density areas in West Malaysia rather than in East Malaysia. These urban centres have existed since the colonial period and have grown over the years. At the micro level, the *mukim* that represent large cities in each metropolitan region (*Kuala Lumpur, Bandaraya Georgetown*, and *Bandar Johor Bahru*) experienced slow or negative growth and declining population shares throughout the 1980-2010 period. *Kuala Lumpur*, the national capital and most urbanised area in Malaysia, has lost some of its attraction since 1980. Since *Kuala Lumpur* is confined to a limited physical space, further expansion beyond its border is absorbed by the surrounding areas. The deconcentration of urban development and urban sprawl are major outcomes of this situation. As large cities rapidly grow, the overflowing development hastens the suburbanisation process in adjacent areas.

Other factors include the establishment of many new townships, major institutions, and industrial centres far away from large cities.

Third, the segregation of ethnic groups by settlement type still existed in 2010. Despite the high population growth of the ethnic majority in most settlement types, ethnic minorities still had the largest population shares in larger settlements. In contrast, the ethnic majority maintained the largest proportion in smaller towns and rural areas. This is an end result of the 'divide and rule' strategy previously imposed in the colonial period to strengthen the colonizer's cause and exploit the country, although the Malaysian government has implemented various policies to integrate ethnic groups.

Fourth, urban-rural segregation can also be seen from a socio-economic perspective. High- and middle-paid workers are more dominant in cities as they are engaged in services or tertiary industry. Low-paid or skilled workers more commonly work in primary industry (agricultural and farming sectors) in rural areas. For secondary industry, there is more construction and manufacturing activity in cities, mainly due to industrialisation and growing urban development. These socio-economic characteristics are known to be the main drivers of population concentration and deconcentration in Malaysia.

Fifth, in relation to differential urbanisation theory, Malaysia may have been in the final stage of urbanisation, *APC*, since 1980, as rapid population growth in capital metropolitan suburban areas has offset a decline in the capital metropolitan core. As the capital metropolitan core area (*Kuala Lumpur*) is saturated with economic and urban development, this has had impacts on agglomeration diseconomies and results in decentralisation towards capital metropolitan suburban areas. Further, the population in all settlement types shows a typical age distribution, with young adults the largest age group. Furthermore, there is no clear sign that Malaysia will move towards the next urbanisation stage (i.e., the polarisation reversal stage) as hypothesized in the theory because the capital metropolitan area contributes to growth and maintain their dominance in terms of relative population compared to smaller cities (regional metropolitan areas, intermediate-sized cities, and small towns villages). However, the real test of this theory will come when examining migration because the theory emphasises net migration change rather than overall population change, and the high population growth in capital metropolitan suburban areas may be influenced by natural population increase instead of migration. This is covered in the next chapter.

Chapter 5

Internal migration and urbanisation in Malaysia, 1980-2010

5.1 Introduction

The previous chapter examined population growth patterns by settlement type over recent decades. The cause of this growth was a contribution of natural population increase and net inward migration. The data do not allow for a detailed examination of natural increase or net migration by settlement type. They do, however, permit a study of internal migration flows. Migration is both an important factor in local population change and a key element in differential urbanisation theory to distinguish between different stages in the urbanisation process. The next section, Section 5.2, provides an overview of the available data and the methods used in this chapter. Section 5.3 then examines internal migration through observation of total migration flows (in-migration and out-migration) and flows between origin and destination. The results show Malaysia experienced rapid rural-urban migration for all settlement types since the period from 1975-1980 due to high in-migration of migrants from lower-level settlements and rural areas. However, the migration pattern changed towards urban-urban migration in the third period (1995-2000) in capital metropolitan areas. On the other hand, rural-urban migration was still high in other cities throughout all periods. Section 5.4 extends the analysis in Section 5.3 by incorporating the age structure of the migration flows. Surprisingly, there was positive growth of in- and out-migrants aged 45 and over in most settlement types. In contrast, there was negative growth of in- and out-migrants aged below 45. Generally young adults have the highest mobility, which then slowly declines with increasing age, sometimes increasing again for those with young children and of retirement age (Bernard et al., 2014). However, despite the large positive growth of in-migrants aged 45, the change in terms of numbers was relatively small. In contrast, negative growth of in-migrants aged below 45 resulted in a significant decline in numbers. The next section, Section 5.5, compares both in-migration and out-migration through a net migration analysis. This section is the core of this chapter as it identifies the urbanisation process as portrayed in differential urbanisation theory. The results of this chapter disagree with the claim made in the previous chapter that Malaysia experienced the final urbanisation stage, the APC stage, since 1980 based on total population and population change alone. This chapter identifies that the country experienced a shift from the IPC stage to the APC stage by 2000 based on the transition of the migration process from rural-urban to urban-urban migration in the largest city, the capital metropolitan area. Finally, this chapter explores the contribution of net in-migration to population growth and the total population in Malaysia. It appears that net in-migration has a clear influence on metropolitan cities. In contrast, net migration in smaller settlements has little or zero influence on population growth and the overall population.

5.2 Data and methods

5.2.1 Data

The main data used for this chapter is the migration matrix of population in the previous place of origin five years ago and current residence by total and age groups at District level (1975-80, 1986-91, 1995-2000, and 2005-2010) which are obtained from the Department of Statistics Malaysia. Table 5.1 shows an example of the migration matrix set:

Table 5. 1: Example of migration matrix of origin and destination population by District.

Origin (5 years ago)	Destination							
	District A	District B	District C	District D	District E	District F	District G	District H
District A	50	150	25	50	30	20	15	15
District B	100	200	50	25	20	30	10	10
District C	75	50	75	150	25	25	10	15
District D	125	50	100	225	25	25	15	10
District E	50	40	75	50	75	150	25	25
District F	30	80	125	50	125	250	25	25
District G	75	70	50	70	75	50	75	200
District H	75	80	50	80	125	50	125	250

5.2.2 Methods

In order to apply the Differential Urbanisation Theory approach, the Districts are first grouped into each settlement type (See Table 5.2 – 5.3)

Table 5. 2: Example of aggregation of District into each settlement type 1

Origin (5 years ago)		Destination							
		District A	District B	District C	District D	District E	District F	District G	District H
		Large cities	Large cities	Medium cities	Medium cities	Small cities	Small cities	Rural areas	Rural areas
District A	Large cities	50	150	25	50	30	20	15	15
District B	Large cities	100	200	50	25	20	30	10	10
District C	Medium cities	75	50	75	150	25	25	10	15
District D	Medium cities	25	50	100	225	25	25	15	10
District E	Small cities	60	50	75	50	75	150	25	25
District F	Small cities	40	100	25	50	125	250	25	25
District G	Rural areas	75	70	60	50	75	50	75	200
District H	Rural areas	75	80	40	100	25	50	125	250

Table 5. 3: Example of aggregation of District into each settlement type 2

Origin (5 years ago)	Destination			
	Large cities	Medium cities	Small cities	Rural areas
Large cities	500	150	100	50
Medium cities	200	550	100	50
Small cities	250	200	600	100
Rural areas	300	250	200	650

Once all Districts have been aggregated into each settlement type, Section 5.3.1 will first examine total migration flows by observing the number of in-migrants and out-migrant. The result in this section will identify which migration is more dominant. For example, if the number of out-migrant is greater than in-migrant, it means more people are moving out of the settlement. In contrast, if the number of in-migrants is greater than out-migrant, then more people are moving into the settlement. Table 5.4 shows the example of the calculation of total in-migrant and out-migrant (exclude within migration flow).

Table 5. 4: Example of total in-migrant and out-migrant calculation (exclude within migration flow)

Origin (5 years ago)	Destination				**Total out-migrant**
	Large cities	Medium cities	Small cities	Rural areas	
Large cities	0	150	100	50	**300**
Medium cities	200	0	100	50	**350**
Small cities	250	200	0	100	**550**
Rural areas	300	250	200	0	**750**
Total in-migrant	**750**	**600**	**400**	**200**	

The next section, Section 5.3.2 examines migration flows between the place of origin and destination. This section allows a further observation on migration pattern where it is rural-urban, urban-rural, urban-urban or rural-rural migration. The result will be explained in terms of the percentage of in-migrants in each settlement type (Table 5.5). The equation is as follow:

$$\%\,Inmigrant\ at\ destination = \frac{\begin{array}{c}Number\ of\ inmigrant\\ from\ different\ origin\\ at\ destination\end{array}}{\begin{array}{c}Sum\ of\ inmigrant\\ from\ different\ origin\\ at\ destination\end{array}} \times 100$$

Table 5. 5: Example of in-migration at destination percentage (exclude within district migration)

Origin (5 years ago)	Destination			
	Large cities	Medium cities	Small cities	Rural areas
Large cities	0	**25**	**25**	**25**
Medium cities	**26.7**	0	**25**	**25**
Small cities	**33.3**	**33.3**	0	**50**
Rural areas	**40**	**41.7**	**50**	0
Total (%)	100	100	100	100

The next step is to incorporate the age group into the analysis which is explained in Section 5.4.1 and Section 5.4.2. These sections use the same method as the previous sections (See Table 5.3 and Table 5.4). For the sake of clarity, ages have been groups by stage in life-course: children/ adolescents (age 0-14), young adult (age 15-29), middle-aged adult (age 30-44), mature adult (age 45-59), and elderly (age 60 and over).

The next step to examine the net migration flows for 1980-1991, 1991-2000 and 2000-2010 in Section 5.5. However, since migration data were recorded on a five-year basis (1975-80, 1986-91, 1995-2000, and 2005-2010), there are no official data on migration for the first five years of each decade (1980-1986, 1991-2005, and 2000-2005). Instead, these missing values are estimated by linear interpolation:

$$NM^3 = \left[\frac{(NM^1 \pm NM^2)}{2}\right] \pm NM^2$$

Where NM^1 is the net migration of the first five year period, NM^2 is the net migration of the second five year period, and NM^3 is the net migration of a new time period.

For example, the equation for net migration of the core city in capital metropolitan for the period of 1980-1991:

$$NM^{1980-91} = \left[\frac{(NM^{1975-80} \pm NM^{1986-91})}{2}\right] \pm NM^{1986-91}$$

$$NM^{1980-91} = \left[\frac{(105{,}576 - 41{,}661)}{2}\right] - 41{,}661$$

$$NM^{1980-91} = -9{,}704$$

Finally, Section 5.6 examines the contribution of net migration to population change and to total population in Malaysia.

$$Net\ migration\ as\ a\ \%\ of\ population\ change = \frac{NM^{t,t+n}}{P^{t,t+n}} \times 100$$

Where $NM^{t,t+n}$ is the net migration of a certain time period and $P^{t,t+n}$ is the population change of a certain time period

$$Net\ migration \text{ as a \% of } total\ population = \frac{NM^{t,t+n}}{AP^{t+n}} \times 100$$

Where $NM^{t,t+n}$ is the net migration of a certain time period and $AP^{t,t+n}$ is the mid-year total population of a certain time period

5.3 Internal migration flows by settlement type

5.3.1 Total flows

Net migration flows are the balance of inflows and outflows in an area. Therefore, before considering net migration, attention must first be paid to observed changes over time in the size of inflows and outflows for each settlement type. Figure 5.1 includes the number of internal migrants with an unknown origin but a known destination. Therefore, the figure does not fully capture the total number of out-migrants for each settlement type. For analysis purposes, it is essential that the distribution of the origins of these migrants is similar to that of migrants with known origins.

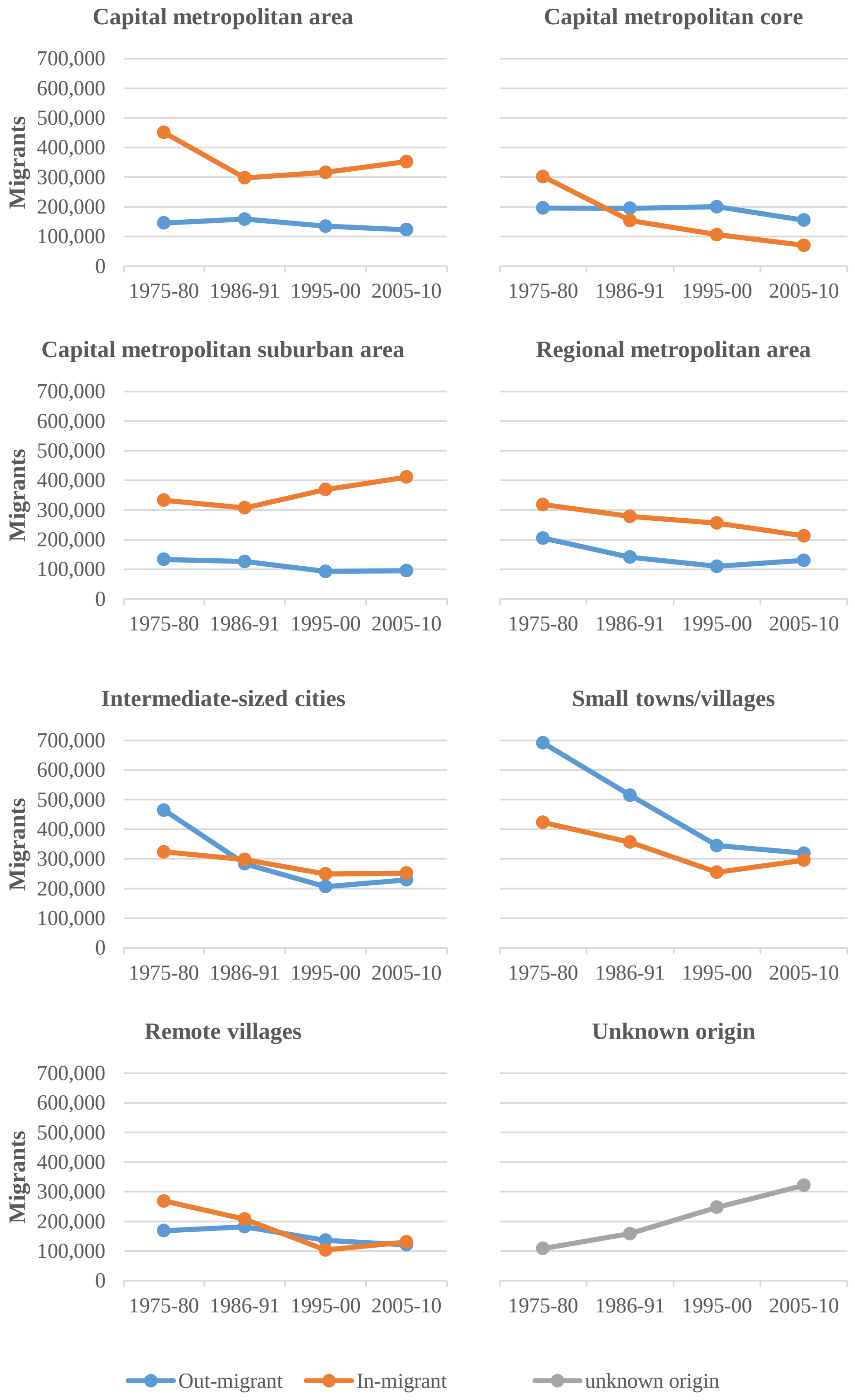


Figure 5. 1: Total number of in-migrants and out-migrants for each settlement and unknown origin out-migrants (excluding those who migrated to the same settlement type)

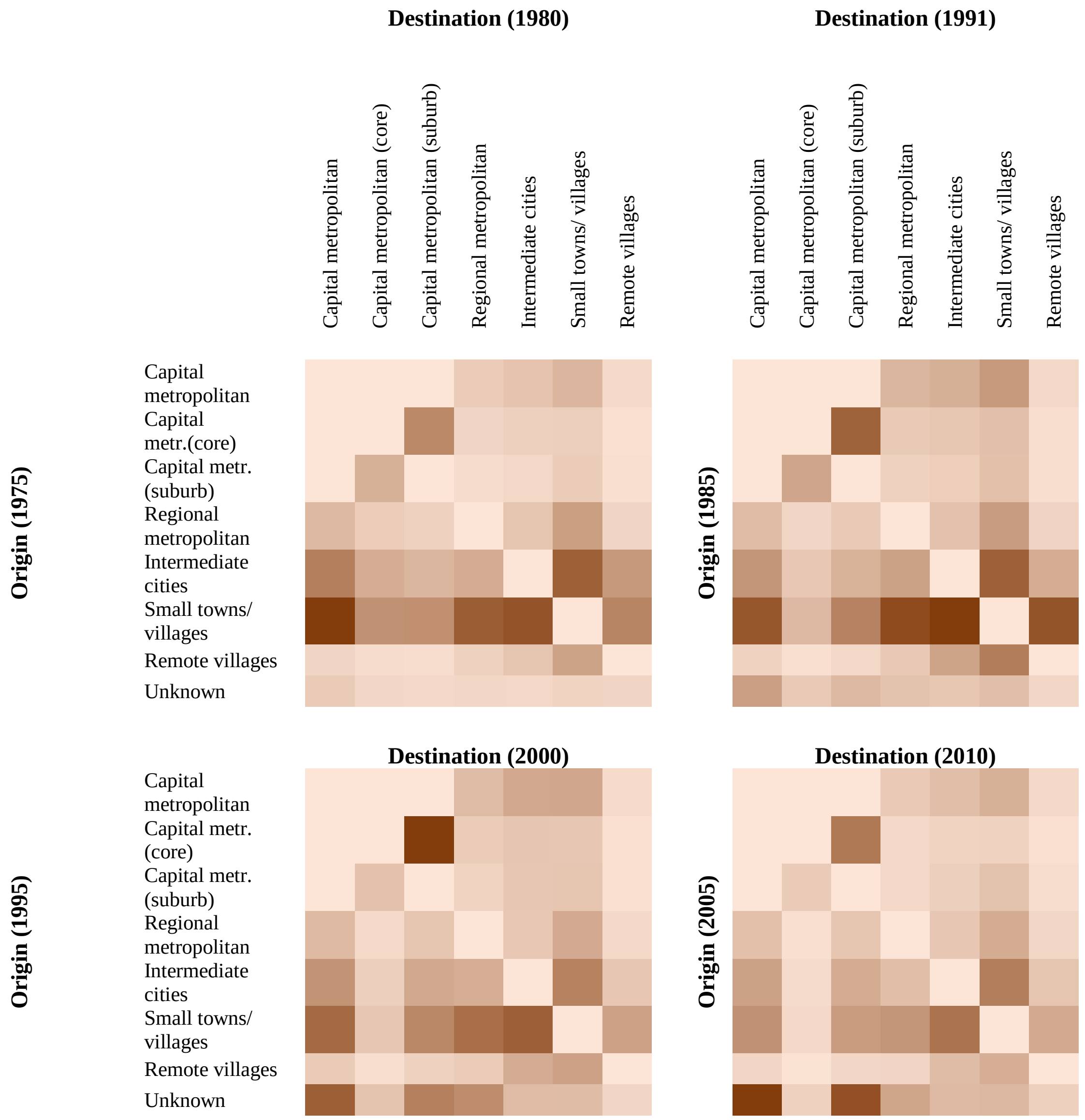
Figure 5.1 shows a major difference for capital metropolitan and regional metropolitan areas display in the number of in-migrants and out-migrants; there are more in-migrants than out-migrants. In other words, more people are migrating into metropolitan cities than moving out of them. However, the capital metropolitan core and suburban areas display an inverse migration pattern; there are more out-migrants in the core and more in-migrants in suburban areas. The contradictory migration patterns in these settlement types indicate that capital metropolitan suburban areas experienced a large concentration of migrants while capital metropolitan cores experienced the opposite. This helps explain the previous chapter's results that indicated the capital metropolitan core had lost its primacy, causing more people to be attracted to the suburban areas since 1980. Rapid suburbanisation caused more people to concentrate in suburban areas than in city cores. Other factors include high costs and maintenance in core cities and the mushrooming of new large-scale urban development (e.g., universities, new townships, and industries) beyond core city boundaries.

Another important finding is that there are more out-migrants for small towns/villages than in-migrants for all periods. Generally, large out-migration flows from this settlement type are due to people's motivation to seek better economic and social opportunities in larger cities. The decision to migrate from a lower-level settlement to a larger city highly depends on income, and living cost differential arises from the spatial segmentation of labour and capital market across regions (Saracoglu & Roe, 2004).

Finally, there was no major difference between the number of in-migrants and out-migrants in intermediate-sized cities and remote villages, except during the initial period of 1975-1980. Higher out-migration from intermediate-sized cities was possibly due to migration towards the nearest or larger cities (capital or regional metropolitan areas). On the contrary, higher in-migration in remote villages was probably due to migrants from neighbouring or nearby small towns/villages.

5.3.2 Flows between origin and destination

As well as considering the aggregated number of migrants, it is important to also examine their origin and destination. This analysis is essential to observe the types of migration flows involved (e.g., rural-urban, urban-rural, urban-urban, and rural-rural). Figure 5.2 shows a heat map of origin-destination flow in Malaysia for four periods (1975-1980, 1985-1986, 1995-2000 and 2005-2010).



Note: Within migration flow is excluded
Details of the origin-destination flow are shown in Appendices section in A2.

Figure 5. 2: Origin-destination flow heat map

Based on Figure 5.2, migrants originating from small towns/villages comprised the largest in all destination settlements in the initial period of 1975-1980. Large out-migration flows from small towns/villages to larger settlements indicate rapid rural-urban (villages to larger cities) as well as urban-urban migration (small towns to larger cities). However, it is difficult to distinguish which migration type is more dominant (rural-urban or urban-urban) because small towns/villages are a combination of small towns and the surrounding villages and rural areas. As mentioned in Chapter 3, this is due to modifiable areal unit problem where the spatial units (districts) used to classify this settlement type is larger than the actual settlement size, hence covering both small cities and the surrounding rural areas. Nevertheless, rural-urban migration may be dominant because small towns/villages were initially established as villages but then evolved into small towns over time. Another important finding is the early deconcentration of population originated from the capital metropolitan core into capital metropolitan suburban areas.

In the second period of 1986-1991, the trends are similar to the previous period whereby most out-migration flows into larger settlements were from small towns/villages. Furthermore, migrants who out-migrated from the capital metropolitan core to capital metropolitan suburban areas continued to have strong shares, indicating a continuous suburbanisation process in capital metropolitan areas.

A notable result during the third period of 1995-2000 is that migrants from unknown origins significantly increased in capital metropolitan and regional metropolitan areas. There is no clear explanation for this because the existing data does not capture the migrants' origins. Moreover, migrants originating from capital metropolitan cores exceeded migrants from small towns/villages in capital metropolitan suburban areas. This result shows a transition in the migration process from rural-urban migration to urban-urban migration during this period. The transition process relates to the changing of urbanisation stages as hypothesized in differential urbanisation theory. Note the claim in Chapter 4 that Malaysia has been in the APC stage since 1980 based on total population and population change alone. However, the results in this section disagree with this claim whereby the country experienced a transition from the Intermediate Primate City (IPC) stage to the APC stage by 2000 based on the migration transition process.

In the final period of 2005-2010, there was a substantial increase of migrants from unknown origin in capital metropolitan areas, with the majority of them residing in suburban areas rather than core cities. These migrants possibly originated in the capital metropolitan core instead of lower-level settlements based on the recent shift in migration patterns. Finally, there remained a large number of migrants migrating from the capital metropolitan core to capital metropolitan suburban areas. This is due to the continuous rapid suburbanisation process in capital metropolitan suburban areas from the mushrooming of new large-scale urban development (e.g., universities, new townships, and industries) beyond the core city boundaries (Abdullah, 2003).

Overall, Malaysia experienced rapid rural-urban migration beginning in the first period due large in-migration from lower-level settlements and rural areas to larger settlements. Most destinations appeared to be in settlements located at the West Coast Peninsular (See A3 in Appendices). One of the main reason for this is the economic transition from agricultural-based towards manufacturing industries in the early 1980s, which attracted large numbers of migrants seeking employment in cities (Mahari, 2011). However, rural-urban migration was replaced by urban-urban migration beginning in the third period of 1995-2000 when the percentage of migrants coming from the capital metropolitan core exceeded the percentage coming from small towns/villages in capital metropolitan suburban areas. These unknown in-migrants in capital metropolitan areas possibly came from core cities rather than lower-level settlements based on a shift of migration patterns. In relation to differential urbanisation theory, the country may have shifted from the IPC stage to the APC stage in 2000, based on the migration transition process. To confirm this, Section 5.5 analyses net migration.

The next section expands this section by incorporating age groups into the analysis.

5.4 Internal migration flows by age group and settlement type

Migration is a selective process, with different age groups commonly exhibiting different migration patterns (Bernard et al., 2014). To examine differences in migration flows for age groups, Section 5.4.1 reviews total flows, and Section 5.4.1 examines flows between origin and destination settlement.

5.4.1 Total flows

This section explains total flows by first reviewing in-migration flows in Figures 5.3 and 5.4, followed by out-migration flows in Figures 5.5 and 5.6

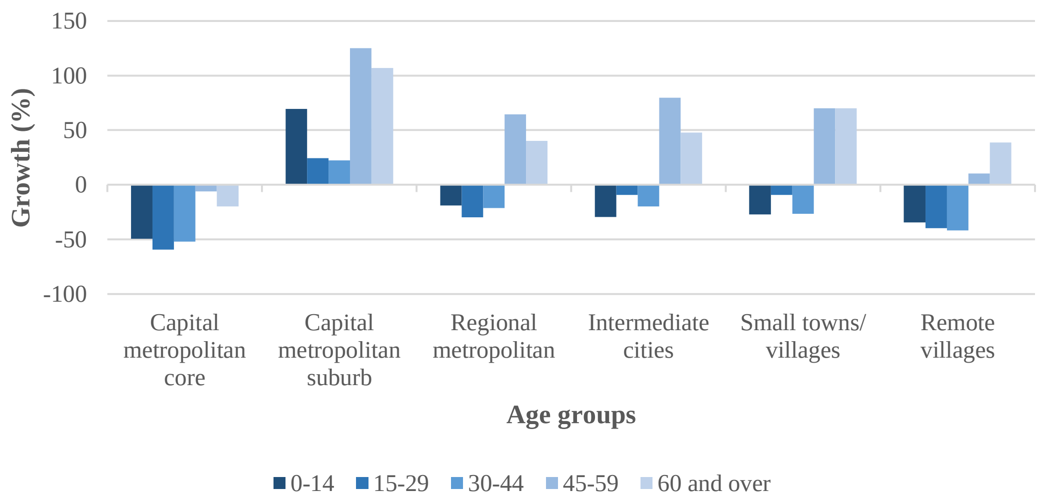


Figure 5. 3: Percentage growth of in-migrants by age group and settlement type, 1980-2010

There are two main findings in Figure 5.3 First, there was positive growth of in-migrants of all age groups in capital metropolitan suburban areas. In contrast, the capital metropolitan core experienced the opposite. This is expected because of the continuous rapid suburbanisation process in capital metropolitan suburban areas since 1980, which attracted migrants of all ages. Second, there was positive growth of in-migrants aged 45 and over in other settlement types. In contrast, there was negative growth of in-migrants aged below 45. This result is surprising because generally young adults (aged 15-29) have the highest mobility, which then slowly declines with increasing age, sometimes increasing again for those with young children and of retirement age (Bernard et al., 2014). To further examine this, Figure 5.4 displays the numbers of in-migrants by age group.

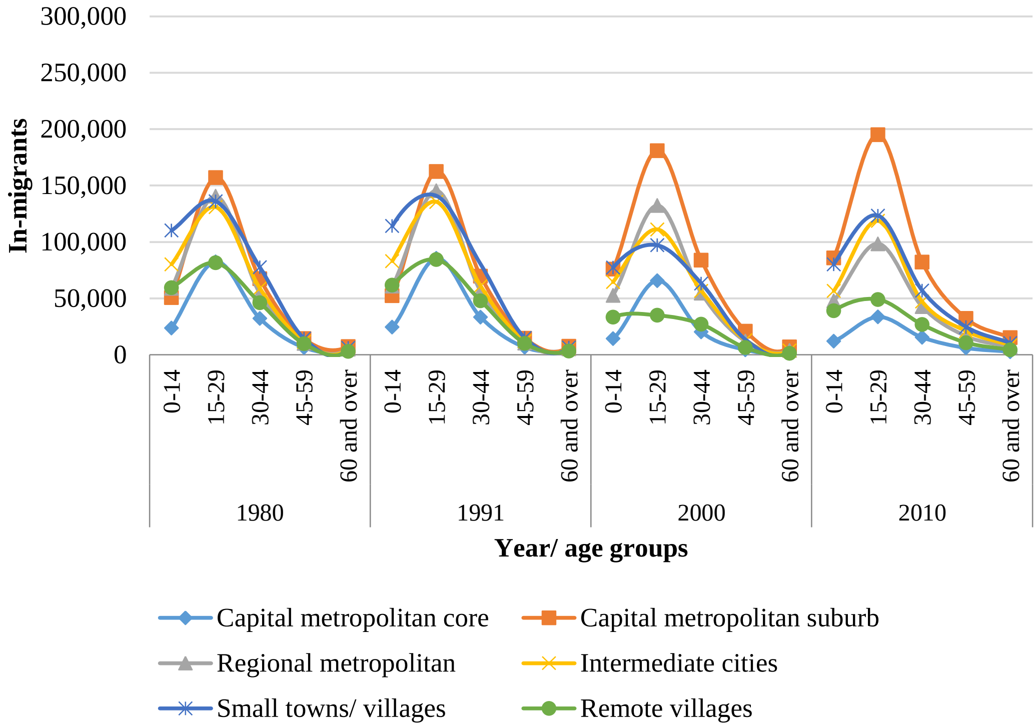


Figure 5. 4: Number of in-migrants by age group and settlement type, 1980-2010

The numbers of in-migrants in Figure 5.4 tell a similar story as the growth results in Figure 5.3. From 1980-2010, the number of in-migrants of all age groups in capital metropolitan suburban areas gradually increased. In contrast, capital metropolitan core experienced the opposite. Despite the large positive growth of in-migrants aged 45 and over in most settlements, the change in terms of numbers was relatively small. In contrast, negative growth of in-migrants aged below 45 resulted in a significant decline in numbers.

All settlements display a typical age distribution, where young adults (age 15-29) have represent the largest number of in-migrants. Young adults generally consist of people leaving school, beginning higher education, entering the labour force, forming unions, and having children. People typically graduate from high school at 17 in Malaysia while the typical age for university graduation is 20-25. Not everyone has the privilege to enter higher education in university or college, and most people migrate to search for jobs after they graduate high school. This is supported in the analysis in Chapter 4 (Section 4.5.4) that found that those with no education and those who had obtained primary education comprised the largest percentages for all settlement types. Furthermore, this can also include young adults who have work commitments in other places that require them to migrate. The same is true for newlywed couples who have committed to stay near or with their partner.

Based on existing studies by UNICEF, following family is the primary reason for internal migration in Malaysia, followed by environmental aspects, career, marital status, and education status (UNESCO, UNDP, IOM, & UN-Habitat, 2012). A wage earners commonly brings with them a partner plus children, hence career-led migration and migration to follow family. For this reason, children/adolescents (aged 0-14) are the second largest group of in-migrants, presumably because each couple has more than two children on average. This group consists of children and adolescents who follow their parents to other cities or migrate for schooling outside their hometowns.

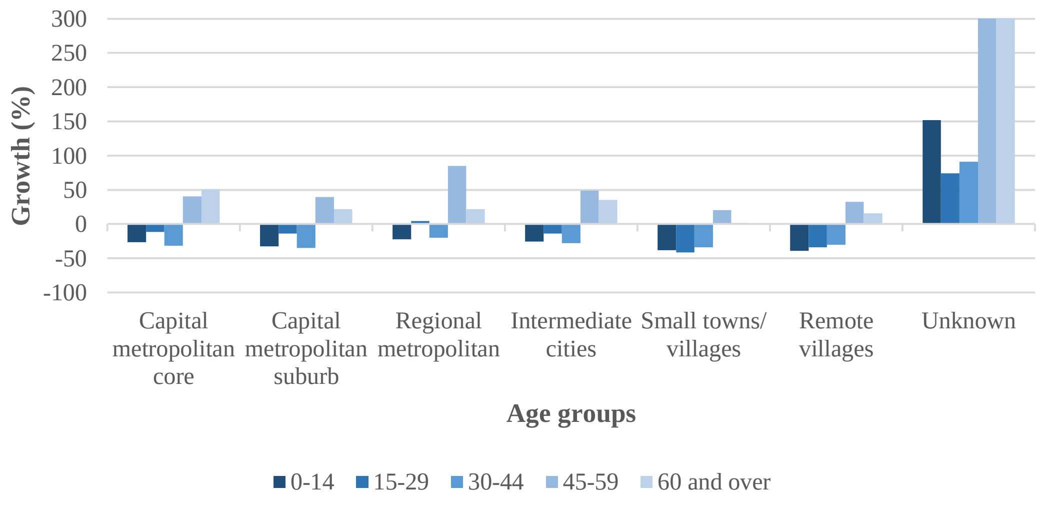


Figure 5. 5: Percentage growth of out-migrants by age group and settlement type, 1980-2010

The out-migration growth in Figure 5.5 is similar to in-migration growth; there was positive growth for out-migrants aged 45 and over in all settlement types. However, this graph includes out-migrants from unknown origins and indicates large positive growth for all age groups. As explained in Section 5.3.2, most of these migrants possibly originated from the capital metropolitan core based on a shift in migration patterns in 1995-2000.

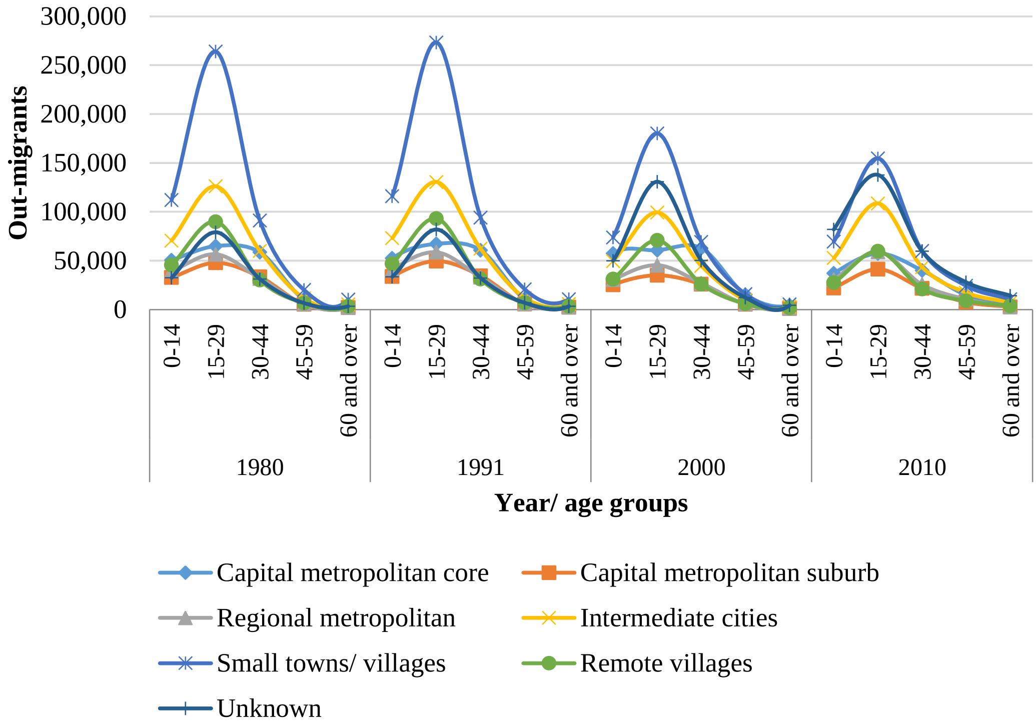
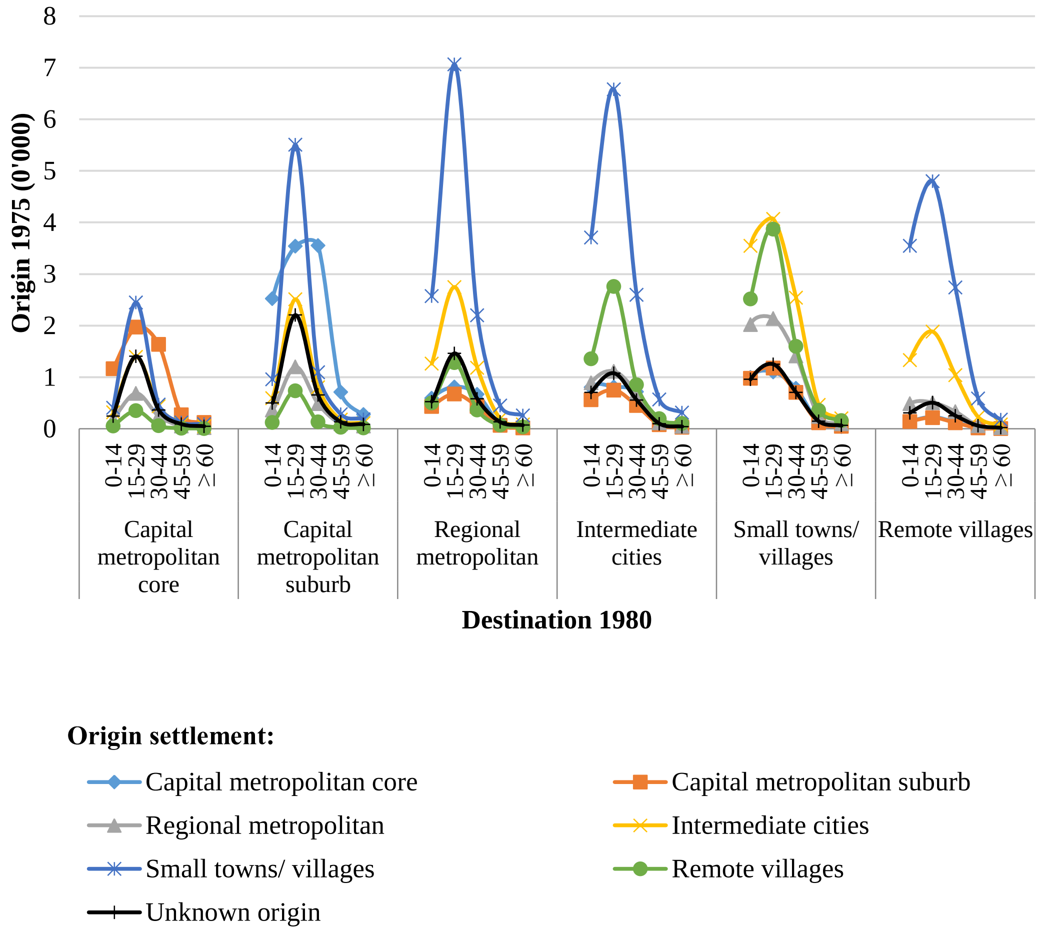


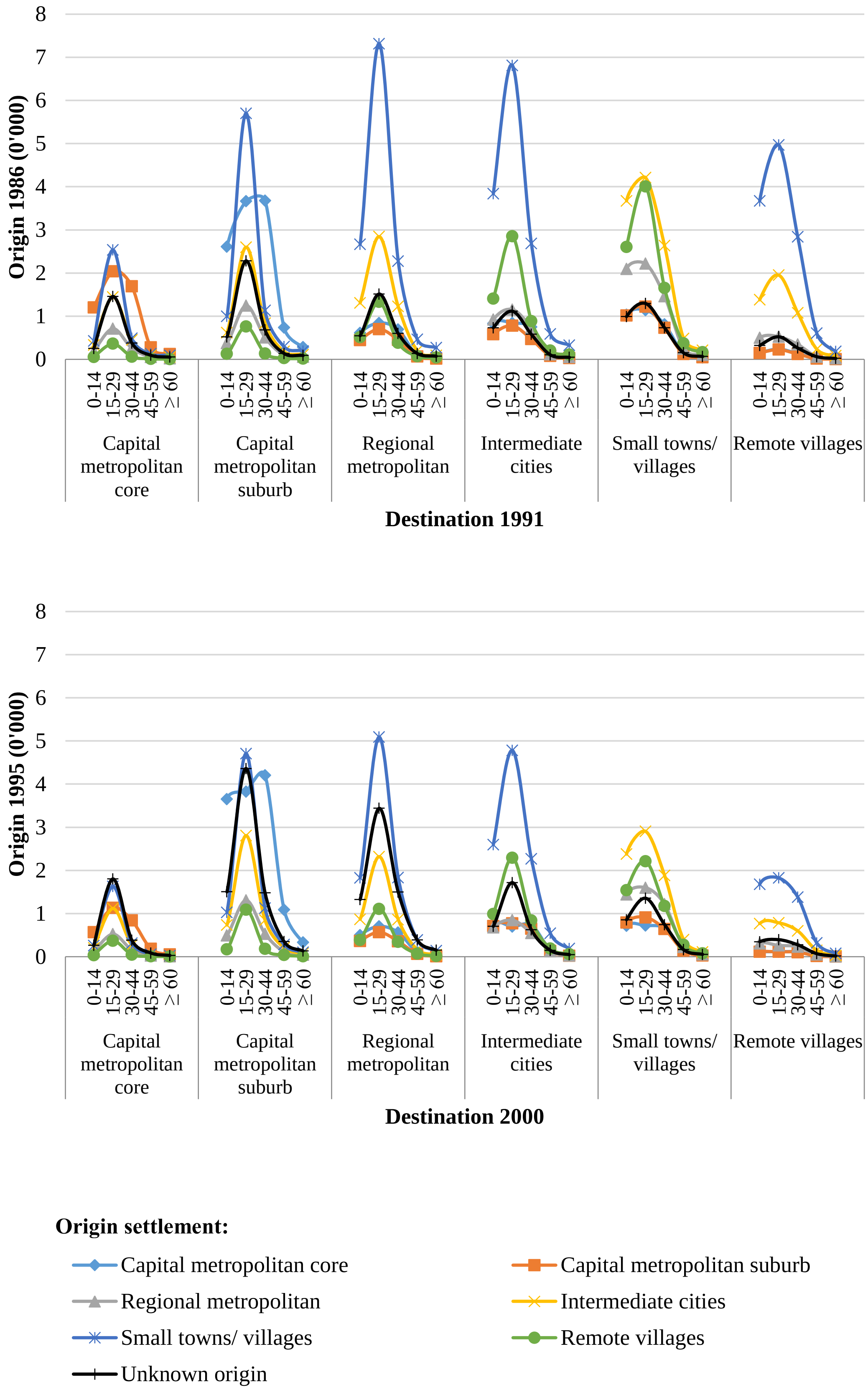
Figure 5. 6: Number of out-migrants by age group and settlement, including the number of out-migrants of unknown origin, 1980-2010

Results in Figure 5.6 for the number of out-migrants has be interpreted similarly as the results in Figure 5.4 for the number of in-migrants; for all settlement types, young adults comprise the largest group. Furthermore, these results agree with the growth results in Figure 5.5. First, the positive growth of out-migrants aged 45 and over has a small effect on the change in numbers. In contrast, the negative growth of out-migrants aged 45 below shows a noticeable decline in numbers for all settlement types across all periods, particularly for small towns/villages. Second, because there was large positive growth for out-migrants of unknown origin for all age, their numbers significantly increased and were almost the same as the number of out-migrants from small towns/villages in 2010.

5.4.2 Flows between origin and destination

The patterns of age-specific migration behaviour outlined above also apply to the population as a whole. As Figure 5.3 and Figure 5.6 illustrate, the same basic pattern also applies to flows between settlements of different types. In particular, the highest numbers for outflow migration, regardless of origin or destination settlement type, were for those aged 15-29 years. However, the numbers and level of mobility varies considerably by origin and destination settlement type (Figures 5.7 and 5.8)





Origin 1986 ('000)
0
1
2
3
4
5
6
7
8
0-14
15-29
30-44
45-59
≥ 60
Capital metropolitan core
Capital metropolitan suburb
Regional metropolitan
Intermediate cities
Small towns/ villages
Remote villages
Destination 1991
Origin 1995 ('000)
Destination 2000
Origin settlement:
Capital metropolitan core
Capital metropolitan suburb
Regional metropolitan
Intermediate cities
Small towns/ villages
Remote villages
Unknown origin

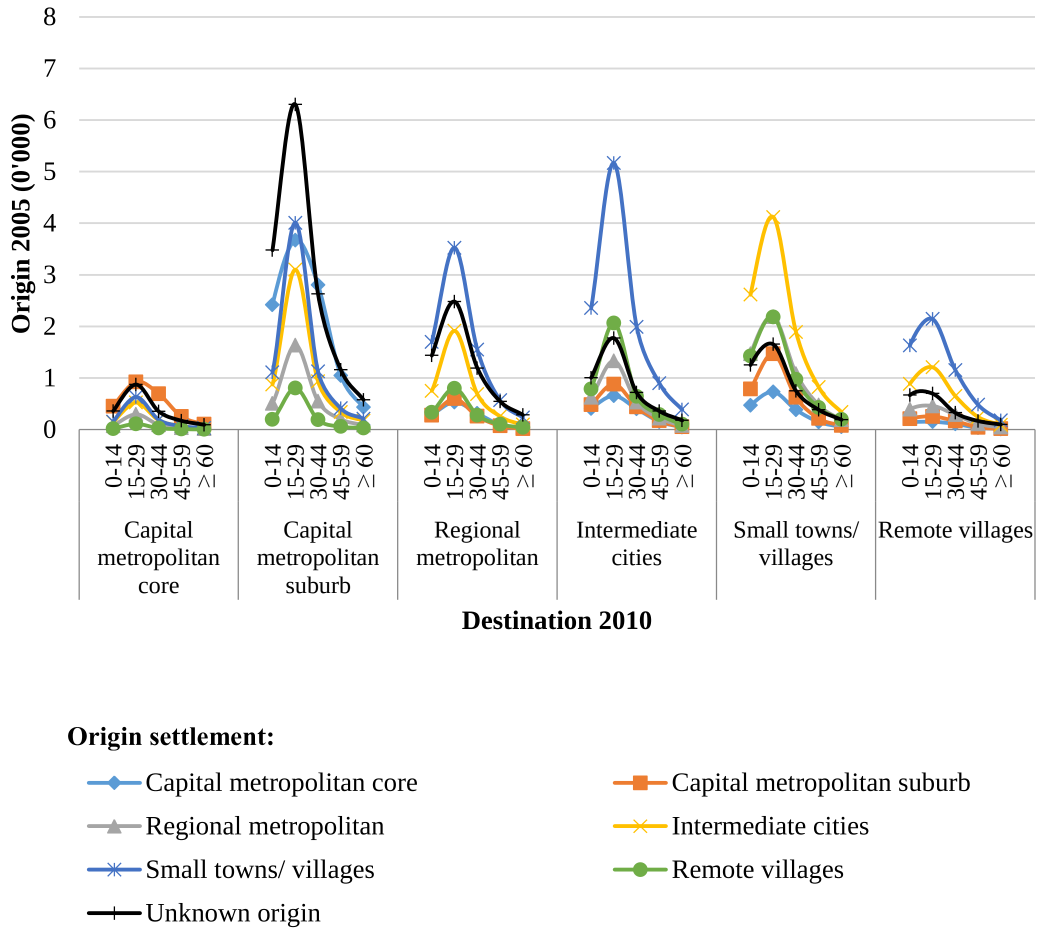


Figure 5. 7: Number of migrants from different origins by destination settlement type and age group

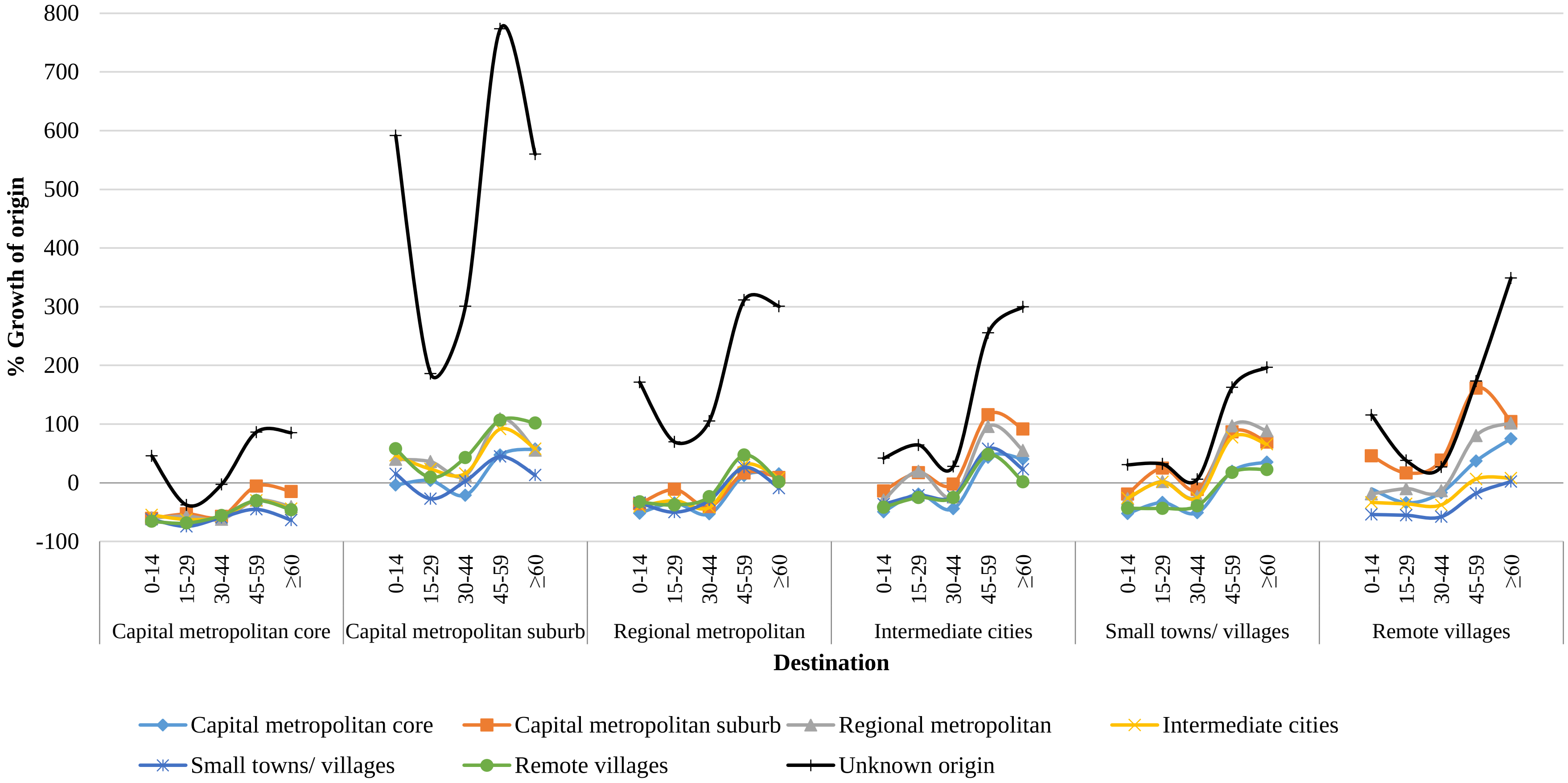


Figure 5. 8: Percentage growth of migrants from different origins by destination settlement type and age group

There are four findings based on Figures 5.7 and 5.8. First, the migration of children/adolescents into capital metropolitan areas (core city and suburban areas) was driven by urban-urban migration throughout all periods: most children/adolescent migrants in the core city came from suburban areas; in contrast, most children/adolescent migrants in suburban areas came from core cities. A possible reason for this is the migration of parents within the vicinity due to changes in residence to be near their workplace or to follow a spouse, which requires children to migrate as well. Further, parent-child migration tends not to occur into smaller settlements or rural areas unless for a specific reason (e.g., job promotion or relocation) because capital metropolitan areas offer better economic and social conditions and opportunities.

Second, regional metropolitan areas and intermediate-sized cities display a common migration pattern: migrants of all age groups (with the largest number of young adults) in these settlements came from smaller towns, typically to seek better economic and social opportunities. This relates to the analysis of socio-economic characteristics and settlement sizes in Chapter 5, where the larger the city, the higher the earnings and the better the jobs that are offered.

Third, most migrants of all age groups in small towns/villages and remote villages came from nearby areas. For example, small towns/villages are near intermediate-sized cities; and remote villages are near small towns/villages. For example, the term 'remote village' is self-explanatory, as such villages are not well connected by major infrastructure and public transportation, hence limiting migration that involves long-distance travel, mainly to larger cities. Furthermore, the rural settlement schemes (e.g., FELDA, DARA, KEJORA) introduced by the Malaysian government during the 1970s to improve the economics of rural communities may be the main reason that people have moved to smaller settlements and rural areas.

Finally, all destination settlement types display similar growth pattern in terms of the origin of migrants; there was major positive growth in the percentage of migrants aged 45 and over compared to those aged 45 below for all settlement types (with those from an unknown origin comprising the largest percentage). Similar to the interpretation in the previous section, despite the high mobility of migrants aged 45 and over, the change in their numbers was relatively smaller than for migrants aged below 45.

5.5 Net migration flows by settlement type

So far, the attention has been on patterns of inflows and outflows. However, overall population change is driven by the net difference between these. Analysis of net migration is important when examining the differential urbanisation theory hypothesis because urbanisation is observed when the net migration in large cities exceeds that in medium and small cities. Further, polarisation reversal occurs when the net migration in medium-sized cities exceeds that in large and small cities, and counterurbanisation occurs when the net migration in small cities exceeds that in large and medium-sized cities. Figure 5.9 shows the net migration to each settlement type in Malaysia for three decades: 1980-1991, 1991-2000, and 2000-2010.

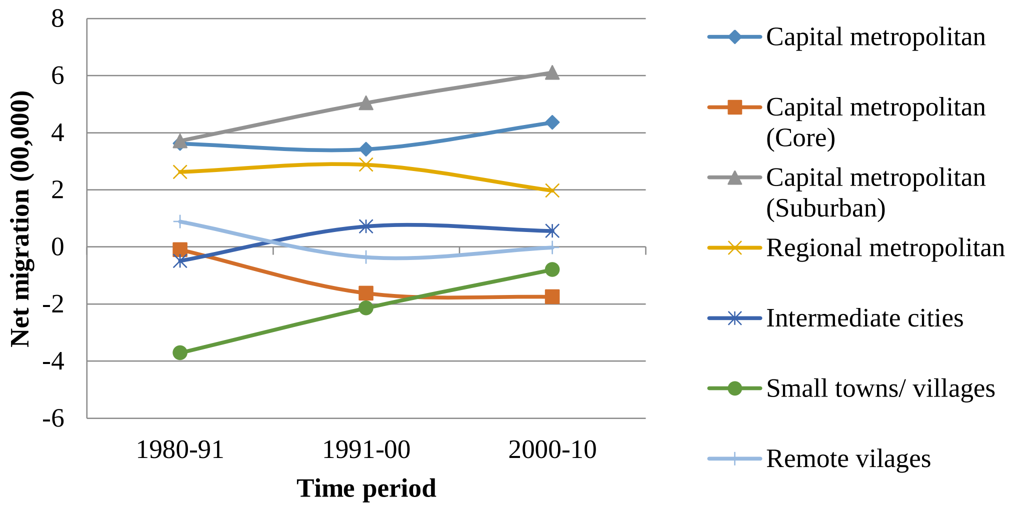


Figure 5. 9: Net migration by settlement type, 1980-2010

There are two main findings based on Figure 5.9. First, metropolitan cities (capital and regional metropolitan areas) maintained large net-inflows throughout the periods. In contrast, the net migration in lower-level settlements (intermediate-sized cities and small towns/villages) was negative at first but then gradually increased over time. The net migration in remote villages, on the other hand, reached a balance by the final period. Second, a large difference in net migration can be seen between the capital metropolitan core and capital metropolitan suburban areas; the capital metropolitan core experienced continuous net-outflows, which is in contrast to capital metropolitan suburban areas, which experienced continuous net-inflows.

In relation to differential urbanisation theory, Malaysia was clearly still in the urbanisation stage due to large net-inflows in the largest city, the capital metropolitan area, during 1980-2010. However, the urbanisation stage comprises three sub-stages: the Early Primate City stage, Intermediate Primate City stage, and Advanced Primate City stage. The results in Section 5.3.2 can be used to identify Malaysia's current stage. The country experienced a shift from the IPC stage to the APC stage by 2000, based on the transition of the migration process from rural-urban to urban-urban migration in capital metropolitan areas.

The APC stage occurs when the core of a primate city is saturated with economic and physical development, resulting in agglomeration diseconomies and decentralisation towards its suburban areas. The capital metropolitan core in Malaysia, *Kuala Lumpur*, is confined by limited physical space and is fully urbanised, so further urban expansion beyond its border will spill over into peripheral suburban areas. Although *Kuala Lumpur* had lost its attraction, the overall capital metropolitan area still maintains its dominance in terms of large net-inflows compared to other settlement sizes, which is consistent with the theory's assumptions.

Another way of thinking about net migration is to examine its relative impact on population change, which is discussed in the next section.

5.6 The relative importance of net migration flows

Population growth arises from the interaction between natural increase and net migration. However, this thesis only examines net internal migration, but not net international migration which is also part and parcel of population growth. Although net international migration is excluded in this analysis, their results would be similar to net internal migration because 1) the size of international migrant stocks (immigrants to and from Malaysia) are relatively small from the overall population – 5.6 percent in 1990, 6.9 percent in 2000 and 8 percent 2010 (UNICEF, 2014), and 2) since internal migration is more dominant than international migration (UNESCO et al., 2012), therefore it should be sufficient to explain the overall story. Obviously the change of international migrants is minimal because the Malaysian government strictly controls the future number of immigrants via work permission and visas.

Besides the relative importance of net migration, this section can also identify the relative importance of natural population increase. For example, if net migration adds 40 percent to population growth, then the remaining 60 percent should be contributed by natural increase. For another example, if net migration does not add to population growth (if it has a negative percentage), then natural increase is entirely responsible for population growth.

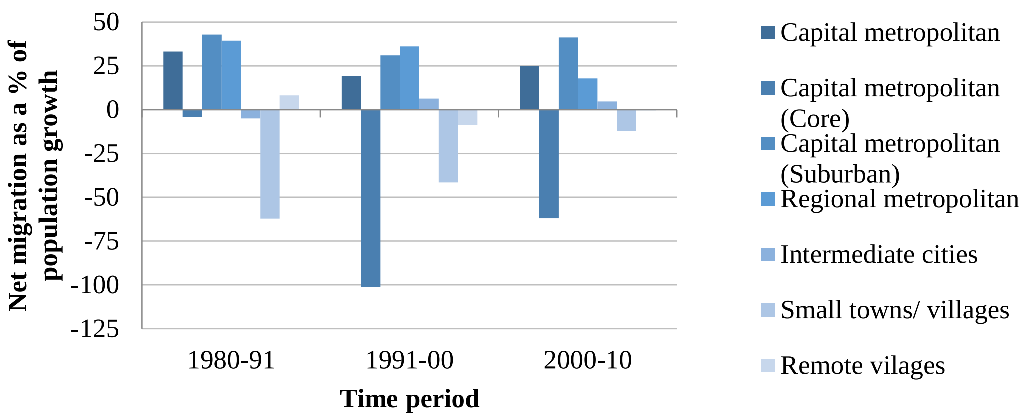


Figure 5. 10: The relative importance of net migration to population growth in Malaysia, 1980-2010

There are four important findings based on Figure 5.10. First, within the 30-year period, net migration patterns in terms of percentages were fairly similar but gradually diminished over time in most settlements (except in capital metropolitan suburban areas), indicating the declining importance of net migration for population growth and the increasing importance of natural population increase. Second, net migration in metropolitan cities (capital and regional metropolitan areas) contributed the largest percentages to population growth compared to other settlement types. Third, net migration did not contribute to population growth in the capital metropolitan core and small towns/villages. Finally, net migration has a minimal influence on population growth in intermediate-sized cities and remote villages. In other words, population growth in metropolitan cities, intermediate-sized cities, and remote villages was mainly due to a natural increase of the population instead of net migration flows. On the other hand, population growth in the capital metropolitan core and small towns/villages was entirely due to natural population increase. In summary, natural population increase was the dominant cause of urbanisation in Malaysia, rather than migration flows.

One of the main reason for this is high fertility levels, primarily after the Second World War period, which resulted in the birth of a large number, who later aged into women of childbearing age. Although fertility in Malaysia has steadily dropped since the 1960s, the number of women of childbearing age increased, resulting in a stable number of births (Hirschman, 1980). This situation relates to the population momentum effect. Population momentum occurs when previous high fertility results in a large proportion of the female population of reproductive age, hence leading to large numbers of births (Keyfitz, 1971). According to Blue and Espenshade (2011), for countries still in the process of demographic transition with a large proportion of women of childbearing age or children that will enter soon enter this period, population momentum can significantly impact population growth. On the contrary, for countries that have completed demographic transition and have a large proportion of ageing individuals and the elderly, along with low fertility, population momentum will not have a significant impact on future growth and the total population will eventually decline (Andreev, Kirill; Kantorová, Vladimíra; Bongaarts, 2013). Besides fertility, a major decline in mortality and increase in life expectancy can also affect the natural increase of population due to improvement in nutrition levels, increasing preventive health programs, and better accessibility to curative medicine (Hirschman, 1980).

Another way of thinking about migration is to consider the size of net migration relative to the size of the settlement population (Figure 5.11)

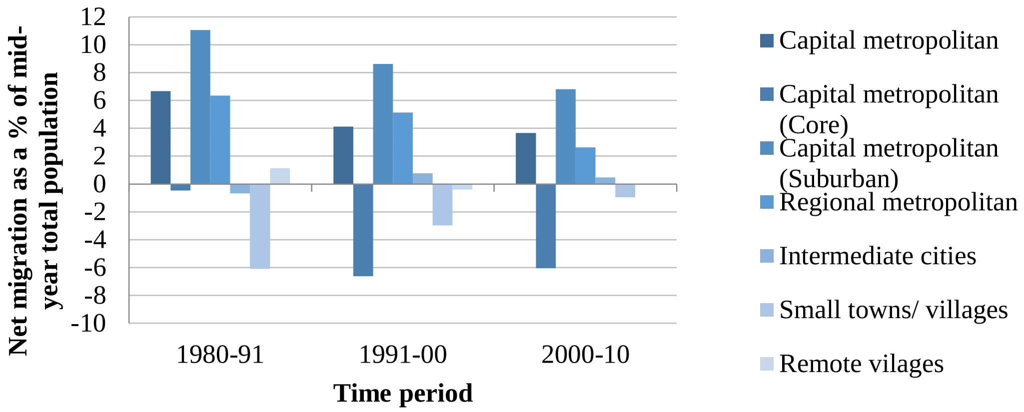


Figure 5. 11: The relative importance of net migration to mid-year total population in Malaysia, 1980-2010

Similar to Figure 5.10, which shows the relative importance of net migration towards the population change, Figure 5.11 shows the relative importance of net migration towards the overall population. The two figures display similar results; within the 30- year period, the percentages follow a similar pattern but diminish over time, with net migration in metropolitan cities (except the capital metropolitan core) making the largest contribution to the mid-year population. In contrast, net migration in other settlement types had little or zero contribution. However, Figure 5.11 highlights the scale of net migration more clearly, indicating it is generally diminishing over time. Over the period from 1980-1991, capital metropolitan suburban areas saw a net-inflow that increased its mid-period population by 12 percent. In 2000-2010, net migration continued to add to the population, but at a slower rate of (7 percent of the mid-year population).

Finally, Figures 5.12 and 5.13 display the spatial distribution of net migration contribution to population change and mid-year total population, respectively. These maps demonstrate that gains and losses were seen in both East and West Malaysia.

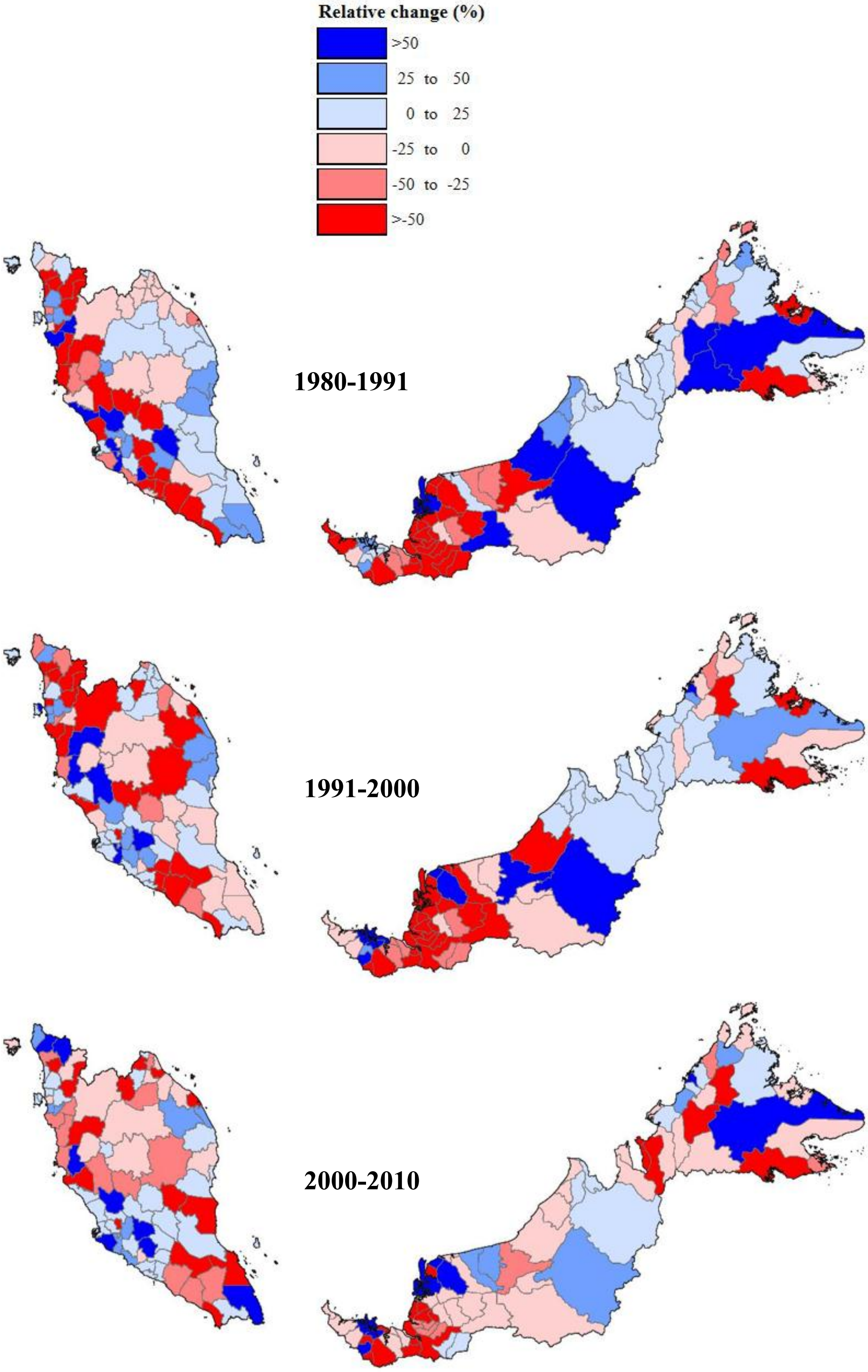


Figure 5. 12: Maps of the relative importance of net migration to population change in Malaysia, 1980-2010

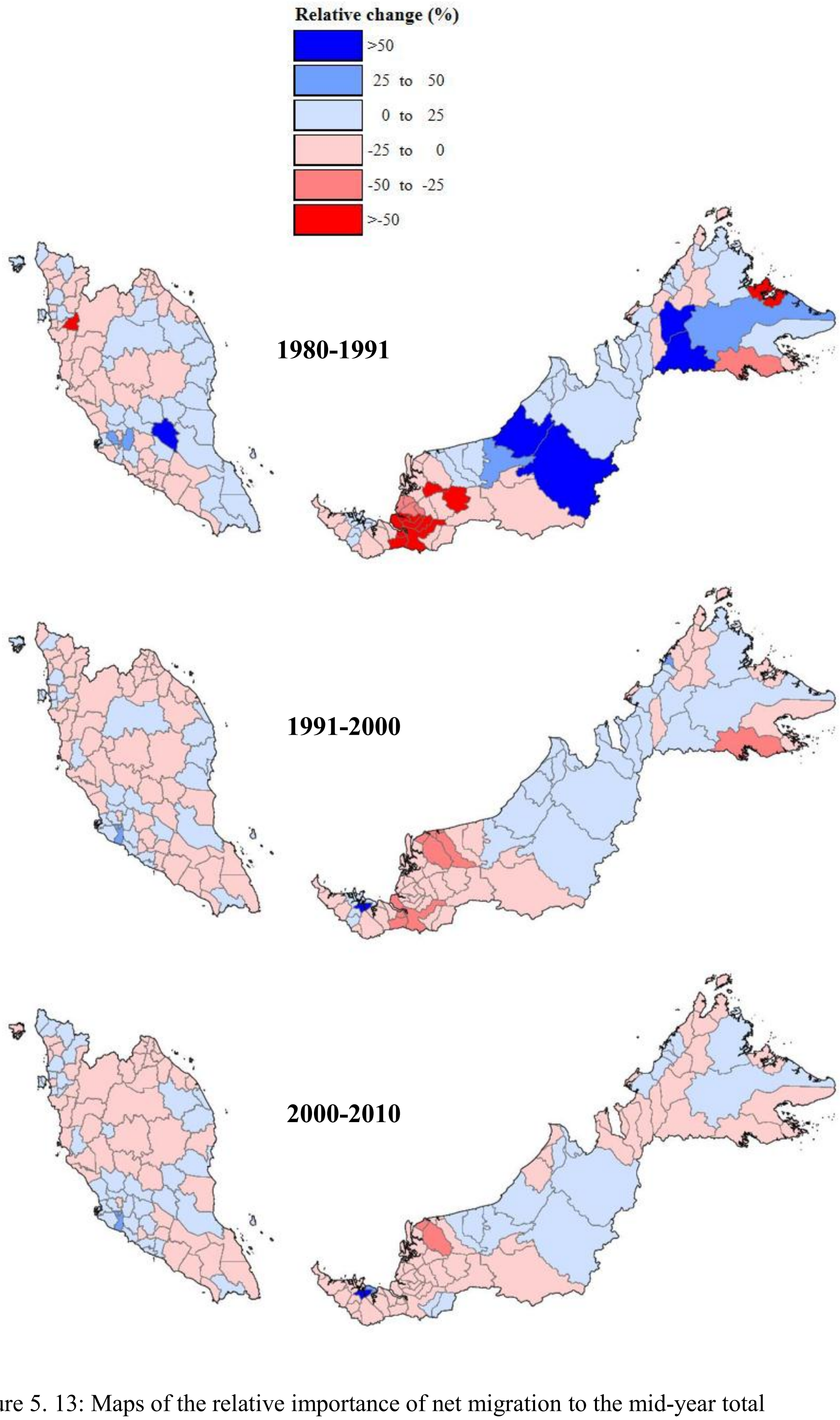


Figure 5. 13: Maps of the relative importance of net migration to the mid-year total population in Malaysia, 1980-2010

5.7 Conclusion

Similar as the contributions in the previous chapter, this chapter extends the analysis of internal migration and urbanisation in Malaysia beyond major urban areas to consider other areas (e.g., intermediate-sized cities, small towns/villages, and remote villages) and analyses urban-rural migration using a classification relevant to differential urbanisation theory.

Overall, Malaysia experienced rapid rural-urban migration during the first period of 1975-1980 due to the large out-migration of migrants from smaller settlements and rural areas to larger settlements. However, in the third period, 1995-2000, the migration pattern changed from rural-urban migration to urban-urban migration when the number of capital metropolitan core in-migrants exceeded small towns/village in-migrants in capital metropolitan suburban areas. In contrast, rural-urban migration remained dominant in other smaller settlement types (regional metropolitan areas and intermediate-sized cities). In terms of the increasing numbers of out-migrants from unknown origins in the capital metropolitan area, there is no clear explanation for this. However, they may have originated within the capital metropolitan area (core or suburban areas) rather than from lower-level settlements based on the changing migration pattern during 1995-2000.

In terms of age groups, in-migrants and out-migrants aged 45 and over had positive growth. In contrast, those aged below 45 displayed negative growth. In other words, those aged 45 and over had higher mobility than those under 45. These results are surprising because generally younger age groups (especially young adults) have the highest mobility, which then slowly declines with increasing age and sometimes increases again for those with young children and of retirement age (Bernard et al., 2014). A possible reason for this is the decline in the number of people of active migrant age (young adults) due to major declines in fertility. This may be true because fertility levels in Malaysia have been declining since the 1960s, with the total fertility rate are at replacement level in 2010. However, despite the high mobility of people aged 45 and over in most settlement types, the change in the number of migrants was relatively small. In contrast, the low mobility of people aged below 45 resulted in a significant decline in the number of migrants.

All settlement types displayed a typical age distribution, with young adults (aged 15-29) comprising the largest group of in-migrants and out-migrants. Young adults generally comprise people leaving school, entering higher education, entering the labour force, forming unions, and having children. People typically graduate from high school at 17 years old in Malaysia while the age at university graduation is between 20-25 years. However, not everyone has the privilege attain higher education in university or college, and most people may migrate to search for a job after they graduate high school. This is supported in the Chapter 4 analysis (Section 4.5.4), which states that those with no education and those who obtained a primary education comprised the largest percentages in all settlement types. Furthermore, this also includes young adults who have work commitments in other places or young newlywed couples who have committed to stay near or with their partner. This relates to existing studies by UNICEF that indicate that following family is the primary reason for internal migration in Malaysia, followed by environmental aspects, career, marital status, and education status (UNESCO et al., 2012). A common reason for this is wage earners bringing with them a partner plus children when they migrate, hence the career-led migration and migration due to following family.

However, migration was not the main cause of urbanisation in Malaysia; instead it was caused by the natural increase of population. The main reason for this is the high fertility levels after the Second World War period, which resulted in a large proportion of woman being of childbearing age. Although fertility in Malaysia has steadily dropped since the 1960s, the number of women of childbearing age continues to increase, resulting in a stable number of births (Hirschman, 1980). This situation may relate to population momentum. Population momentum occurs when previous high fertility results in a large proportion of the female population being of reproductive ages, hence leading to large numbers of births (Keyfitz, 1971). Besides fertility, other factors include a major decline in mortality and increase in life expectancy since at least the 1950s.

In relation to differential urbanisation theory, the results of this chapter disagree with the claim made in the previous chapter that Malaysia experienced the final urbanisation stage, the APC stage, beginning in 1980 based on total population and population change alone. This chapter identifies that the country experienced a shift from the IPC stage to the APC stage by 2000 based on the transition of the migration process from rural-urban to urban-urban migration in the largest city, the capital metropolitan area. APC is the stage in which the capital metropolitan core is saturated with economic and physical development, causing agglomeration diseconomies and decentralisation towards capital metropolitan suburban areas. Further, since the capital metropolitan core, *Kuala Lumpur*, is confined by limited physical space, further urban expansion will go beyond its borders. Although the capital metropolitan core lost its attraction to many, the overall capital metropolitan area maintained its dominance in terms of large net-inflows compared to other settlement sizes, which is consistent with the theory assumption.

Existing urbanisation studies explain migration flows based only on historical context or recent experiences in Malaysia. Furthermore, data on district-level migration flows (which is used to examine change by settlement type) released by the Malaysia Department of Statistics disaggregate these flows only by age. In the absence of information about other attributes of migrants, the next chapter explores the drivers of these migration flows through the application of a series of spatial interaction models, using the socio-economic attributes of the origin and destination areas as proxies for migration push and pull factors.

Chapter 6

Determinants of internal migration in Malaysia, 1980-2010

6.1 Introduction

Following internal migration studies in Chapter 5, this chapter investigates the determinants of migration flows in Malaysia from 1980 to 2010. Generally, there are various determinants of migration, such as demographic, economic, social, cultural, environmental, spatial, and housing factors (Champion, Fotheringham, Rees, Boyle, & Stillwell, 1998). Historically, rural-urban migration in developed countries (particularly during the industrialisation period) was mainly driven by economic opportunities in cities (e.g., jobs in manufacturing industries), which attracted rural communities to migrate to urban areas. However economic factors had less influence on migrations flow when the migration pattern changed from rural-urban to urban-rural migration in the late 19th century (Berry, 1980; Champion, 2003; Fielding, 1982). Instead, migration flows were influenced by a combination of factors (e.g., demographic, environmental, cultural, spatial, and housing factors). The experience of developed countries, on the other hand, may be insufficient to explain migration behaviour in developing or less developed countries due to the complex nature of these countries. For example, rural-urban migration in sub-Saharan Africa was influenced by technological and institutional change instead of industrialisation factors (Fox, 2012). The technological and institutional change introduced by colonizers diminished major problems in African cities (e.g., food shortages and limited access to disease prevention methods), thus attracting more rural-urban migration. Besides technological and institutional factors, environmental and socio-economic factors (e.g., agricultural crises, famine, droughts, and poverty) also played an important part in the migration process (Hosszú, 2009). In terms of cultural factors, men in Kenya are known to be the breadwinners, thus often requiring them to migrate to support their families (Mihermutu, 2011). Nevertheless, there is considerable evidence that economic factors play a major part in rural-urban migration in both developing and less developed countries.

To explain migration flows and identify the determinants of internal migration in Malaysia, a variety of techniques were adopted, and several migration theories were reviewed in Chapter 2. Common migration models include spatial interaction models, binary logit models, multinomial logit models, trend extrapolation models, and microsimulation models. Because the purpose of this chapter is to identify migration determinants and all the data are aggregated at area level, spatial interaction models are preferable to other models (Sections 6.2 to 6.4 discuss this matter in detail). The modelling is tackled in two ways: 1) by modelling total migration flows in Section 6.5 (e.g., net migration, in-migration, and out-migration); and 2) by modelling migration flows between origin and destination in Section 6.6. The reason total migration flows are included is to examine how well the determinants are being captured by the model at an aggregate level. The results of aggregate flow models, however, captures only a few determinants (e.g., the net migration model captures only one determinant). This is because this approach models only aggregate flows to each place, rather than the full set of individual origin-destination flows. Further, according to Rogers (1990), there is no such thing as net migration because migrants only consists of people who in-migrate and out-migrate. On the other hand, the results for spatial interaction models for origin-destination flows in Section 6.6 capture plausible lists of determinants, hence explaining migration flow better than the aggregate models.

6.2 Brief review of migration models

According to Stillwell and Congdon (1991), it is impossible to categorise migration models because they are not ‘mutually exclusive’. There are many different migration models applied in the literature, and it is difficult to find multiple studies that have used similar models (Champion et al., 1998). Nevertheless, it is possible to categorise models that have commonly been used. In the Malaysian context, only a few studies have applied modelling techniques to examine migration. For example, the relationship of migration to fertility (Bach, 1981), career (Chattopadhyay, 1998), ethnic concentration (Chitose, 2001) and income and unemployment (Hussain & Abdullah, 2014).

According to Champion, Fotheringham, Rees, Boyle, and Stillwell (1998), migration can be observed from a macro or micro perspective; the former relates to demographic process or change in a region while the latter relates to the 'outcome of individual's decision-making process'. Migration can also be categorised by the use of cross-sectional data (i.e., migration in one time period) or time-series data (i.e., migration in more than one time period). Based on these reasons, Champion, Fotheringham, Rees, Boyle, and Stillwell (1998) introduced four types of migration models: aggregate (macro) cross-sectional models, disaggregate (micro) cross-sectional models, aggregate (macro) time-series models, and disaggregate (micro) time-series models.

6.2.1 Aggregate (macro) cross-sectional models

Aggregate (macro) cross-sectional models are the most common type of migration models and are commonly referred to as spatial interaction models. The models have been used in many migration studies (see Dennett & Wilson, 2013; Sapiro, 2017; Vobruba, Körner, & Breitenecker, 2016; and Wilson, 2006) as well as in other fields such as economics, town planning, and transportation studies (e.g., modelling traffic flows and market trading). Spatial interaction models (also known as gravity models) use a statistical approach to estimate or predict interaction and flows between populations and regions (Borjas, 1994). The data are compiled for one time period, and the migration flows can be aggregated by total or by cohort. The migration matrix includes total migration flows between origin and destination; it can also be a matrix of migration flow by cohorts (e.g., age groups, ethnicity). The general equation of spatial interaction models is as follows:

$$M_{ij} = f(X_i, X_j, X_{ij})$$

where M_{ij} represents the migration flow between origin and destination; f( ) is a functional form; X_i represents attributes that push migrants from an $origin_i$; X_j represents attributes that attract migrants to a $destination_j$; and X_{ij} represents attributes describing the spatial separation of origin and destination (e.g., distance, crossing the sea).

In general, the outcomes are modelled based on a selected set of explanatory variables (one or more). There are three types of spatial interaction migration models that can be derived based on the general equation: the production or origin-constrained model; the attraction or destination-constrained model; and the doubly constrained model. A detailed description of these models is provided in Section 6.4.2.

6.2.2 Disaggregate (micro) cross-sectional models

The application of disaggregate (micro) cross-sectional models is basically the same as aggregate (macro) cross-sectional models but uses individual migrant data instead of aggregated migrant data. Equivalent to the destination-constrained model in the previous section, a binary logit model can be used to estimate or predict the decision of an individual to migrate from an origin. In contrast, a multinomial logit model is equivalent to the origin-constrained model and can be used to estimate or predict the destination choice of migrants in terms of destination attributes and personal attributes (Champion et al., 1998). Recent examples of the application of a multinomial logit model for migration include Soon (2010), Suryawanshi, Sharma, Saggurti, and Bharat, (2016), and Olowa, (2012). The general equation of disaggregated models is as below:

$$P_{hi} = f(X_i, Z_h)$$

where P_{hi} represents the probability a person or household h in place i will migrate; Xi represents the characteristics of location i where person h lives; and Z_h represents the of attributes person h. The characteristics of location i may include social and economic conditions at the origin. The person or household attributes include age, marital status, income, housing tenure, occupation, and other factors that may affect the decision to migrate.

Disaggregated models' outcomes are more accurate than those of aggregated models because migration involves the decision-making of individual migrants or households instead of decisions by an aggregate group of migrants (Champion et al., 1998). Further, the models are more refined than aggregated models because the explanatory variables (e.g., income, age) can be included as continuous variables instead of aggregating the variables. For example, aggregated models may combine age groups (e.g., 0-14, 15-29, 30-44…) while disaggregated models may include age as continuous a variable (e.g., 0,1,2,3...). However, a shortcoming of these models is the intensive use of data required and few applications in the literature.

6.2.3 Aggregate (macro) time-series models

Aggregate (macro) time-series models are commonly referred to as trend extrapolation models. Trend extrapolation models are used to predict future migration by assuming the migration rate is constant in the future (Smith, Jeff, & David, 2017). Unlike the previous two models, trend extrapolation models cannot explain current or past migration flows and only predict migration flows between origin and destination. The models are especially useful for short-term rather than long-term forecasting because there is a small chance of major change in demographics and the social and economic conditions of a region over a short period (Armstrong, 1985; Champion et al., 1998). However, the application of these models is limited if geographical or administrative boundaries change over time due to lack of time-series data or due to the inapplicability of a model to the new boundaries (Champion et al., 1998). The accuracy of these models is reasonable because they can be based on previous or recent time-period migration data to forecast future migration. One of the most famous extrapolation models is Markov chain analysis. Markov chain analysis is highly dependent on the assumption that the migration matrix remains constant over the forecast period. Examples of studies that have applied Markov chain analysis for migration are Joseph (1975), Rogers (1966), and Zimmermann and Constant (2012). The general equation of the model is shown below:

$$m_{ij} = M_{ij} / P_i$$

where m_{ij} represents the probability that the population originating in region i survives in region j at the end of the time interval, M_{ij} represents surviving migrants from origin i to destination j, and P_i represents the population of i. Future migrants can then be calculated by:

$$P^{t+1} = m_{ij} . P^t$$

where P^{t+1} represents the regional population at time t+1, m_{ij} is the transition probability migration matrix, and P^t represents the population at time t.

6.2.4 Disaggregate (micro) time-series models

Disaggregate (micro) time-series models use micro time-series data and aim to forecast individual migration behaviour under various demographic and socio-economic conditions (Champion et al., 1998). The models are also known as microsimulation models. Microsimulation models are used when aggregate models (e.g., Markov chain analysis) cannot explain factors that affect individual or household behaviour or factors influencing the behaviour are complex (Champion et al., 1998). Early applications of microsimulation models include migration studies of specific ethnic groups over a certain time period using a grid map (Hansell & Clark, 1970; Morrill, 1965; Woods, 1981). A recent application of the model is for long-term population projection by accounting for overseas migration (King, Walker, & Bækgaard, 2002). The model deals with samples (e.g., individual characteristics such as age, sex, marital status, and income) rather than aggregate data (e.g., total population) to avoid aggregation bias and produce more detailed forecasts (Smith et al., 2017). Microsimulation models have long been used to analyse the effect of policy on migration trends, but the application of the model is extremely complex and it is difficult to obtain data (Smith et al., 2017).

6.2.5 Summary

In summary, the accuracy of both cross-sectional and time-series models (aggregated or disaggregated) is reasonable for only short forecast periods. The models are useful for short-term forecasting because there is a small chance of major change in demographics and the social and economic conditions of a region. Furthermore, spatial interaction models (i.e., binary logit models, multinomial logit models, and microsimulation models) are suitable to examine small-area units that are known to have large spatial differentiation. On the contrary, trend extrapolation models are more accurate for large spatial units because the effects of unusual conditions can be averaged out or reduced. Moreover, the application of a trend extrapolation model is limited if geographical or administrative boundary changes over time because there will be more error if the forecasted data are limited to a certain number of regions. On the contrary, spatial interaction models offer more opportunities to examine geographical change. Because small-area units are vulnerable to change, explanatory models are thus appropriate for this condition. Unlike spatial interaction models, trend extrapolation models are cannot explain current or past migration flows due to changes in economic and social conditions and have only been used to forecast total migration flows (or by cohort).

Overall, to select a suitable model for modelling migration, it is important to know the purpose of this analysis and what data is available. If the purpose is to understand how migration is affected by the economic and social conditions of a region, then spatial interaction models (for aggregated data), binary logit models, and multinomial logit models (for disaggregate data) are preferable. On the other hand, the trend extrapolation method is better if such information is not needed or the purpose is to estimate or forecast total migration flows.

Because the purpose of this chapter is to investigate the determinants of migration and all data are aggregated (see Section 6.3 for more detail), spatial interaction models are preferable. The modelling approach is done in two ways: 1) by total migration flows in Section 6.5 (e.g., net migration, in-migration, and out-migration); and 2) by origin-destination flows in Section 6.6. Total migration flows are modelled before the origin-destination flow to examine how well the model captures the determinants at the aggregate level. The next section introduces data for both spatial interaction modelling approaches.

6.3 Data

This chapter uses a combination of datasets obtained from the Department of Statistics Malaysia and IPUMS International. The dataset from Department of Statistics Malaysia consist of aggregated migration flows (e.g. in-migration, out-migration and net-migration) and flows from origin to destination by district level, 1970-2010; while the dataset from IPUMS International consists of socio-economic characteristics of the population by district level, 1970-2000. The reason dataset from IPUMS International is used is that the Department of Statistics Malaysia only published socio-economic data on the population by State level which is not essential for this study. However, datasets from IPUMS International is still considered as official data because they retrieved it from the Department of Statistics Malaysia. Table 6.1 shows the category and sub-category of data from both sources.

Table 6. 1: Source and type of data

Source	Category	Sub-category (District level)	Years available
Department of Statistics Malaysia	Migration flows	Number of in-migrants	1970-2010
		Number of out-migrants	
		Number of net migrants	
		Number of flows taken from origin to destination	
IPUMS International	Population	Number of total population	1970-2000
	Sex	Number of males	
		Number of females	
	Age-group	Number of child/adolescents (aged 0-14)	
		Number of young adults (aged 15-29)	
		Number of middle-aged adult (aged 30-44)	
		Number of mature adults (aged 45-59)	
		Number of elderly (aged 60 and over)	
		Number by working-age group (aged 15-65)	
	Ethnicity	Number of ethnic majority	
		Number of ethnic minority	
	Marital status	Number of single	
		Number of married	
		Number of widowed/ separated	
	Education Attainment	Number of without formal education	
		Number of primary education	
		Number of secondary education	
		Number of tertiary education	
	Employment Status	Number of employed	
		Number of unemployed	
		Number of inactive	
	Types of occupation	Number of low-pay jobs	
		Number of medium-pay jobs	
		Number of high-pay jobs	
	Types of occupation industry	Number of primary industry worker	
		Number of secondary industry worker	
		Number of tertiary industry worker	

There two problems with the IPUMS International dataset: 1) the sample only covers 2 percent from overall District population; and 2) there is no information for 2010. In contrast, migration data from the Department of Statistics Malaysia covers all migration flows in each District for all census years. In order to fit both datasets together, first, all information from IPUMS International dataset need to be converted into percentages. Table 6.2 show an example of adjustment made of the IPUMS International dataset.

Table 6. 2: Example of adjustment of population by sex of District A

Sex	Number of population (a)	Percentage (a) / Total (a) *100
Male	253	50.8
Female	245	49.2
Total	498	100

Once all information has been converted into percentages, then the next step is to prepare and organize the datasets based on the selected modelling techniques.

6.3.1 Aggregate flow models

This modelling approach uses a dataset with one row per census per District. In total, there are 510 number of observations (District) from 1970 until 2010: 119 for 1970-1980, 127 for 1980-1991, 131 for 1991-2000, and 133 for 2000-2010. The reason each period consist of a different number of Districts is due to the changes in geography as time passes. Each row of District consists of 38 columns of variables (Table 6.3).

Table 6. 3: Variables for aggregate migration modelling

Variables	Category	Sub-category
Dependant variables	Migration flows	Number of in-migrants
		Number of out-migrants
		Number of net migrants
Independent variables	Population	Number of total population
	Sex	% of Male
		% of Female
	Age-group	% of Child/adolescents
		% of Young adult
		% of Middle-aged adult
		% of Mature adult
		% of Elderly
		% of Working age group
	Ethnicity	% of Ethnic majority
		% of Ethnic minority
	Marital status	% of Single
		% of Married
		% of Widowed/ Separated
	Education Attainment	% of Without formal education
		% of Primary education
		% of Secondary education
		% of Tertiary education
	Employment status	% of Employed
		% of Unemployed
		% of Inactive
	Types of occupation	% of Low pay jobs
		% of Medium paid jobs
		% of High paid jobs
	Types of occupation industry	% of Primary industry worker
		% of Secondary industry worker
		% of Tertiary industry worker
	Types of settlement (Dummy variables)	Capital metropolitan core
		Capital metropolitan suburb
		Regional metropolitan
		Intermediate cities
		Small towns/ villages
		Remote villages (excluded)
	Years (Dummy variables)	Year 1970-1980 (excluded)
		Year 1980-1991
		Year 1991-2000
		Year 2000-2010

Once the dataset is prepared, the next step is to examine the data distribution to check whether it is more or less symmetrical. This can be checked by calculating skewness or visualising using a histogram. If the variable is highly skewed (e.g. skew more than 1.0), then it should be transformed before inclusion into the model. To transform the variable, first the minimum value of the variable need to be checked whether it is negative or 0. If it is 0, the distribution has to be shifted to the right to make all values positive (e.g. if the minimum value is -30, add a constant value such as 31 to every value, which means the new minimum will be 1). This transformation will leave the shape of the distribution, and its skewness, unchanged. To make the distribution more symmetrical:

- If the skew is positive, then possible transformations tested are log and square-root.
- If the skew is negative, then possible transformations tested are square and cube.

Whichever transformation results in a skew closest to zero is the one used as the dependent variable for regression purposes. If no transformation produces a skew closer to zero than the skew of the original untransformed distribution, then the original version of the variable issued. Table 6.4 shows the data transformation of variables for the aggregate migration model.

Table 6. 4: Data transformation for aggregate migration model

Variables		Skewness	Best transformation	Skewness after transformation
Dependant variables	Number of in-migrants	3.81	natural log	0.14
	Number of out-migrants	4.12	natural log	0.05
	Number of net migrants	2.80	square root (y+94,248)	-1.97
Independent Variables	Total population	4.13	natural log	0.30
	% of Male	0.66	-	0.66
	% of Female	-0.66	-	-0.66
	% of Child/adolescents	0.00	-	0.00
	% of Young adult	0.97	-	0.97
	% of Middle-aged adult	0.28	-	0.28
	% of Mature adult	0.11	-	0.11
	% of Elderly	0.55	-	0.55
	% of Working age group	0.14	-	0.14
	% of Ethnic majority	-0.60	-	-0.60
	% of Ethnic minority	0.60	-	0.60
	% of Single	-0.67	-	-0.67
	% of Married	0.45	-	0.45
	% of Widowed	0.56	-	0.56
	% of Without formal education	-0.10	-	-0.10
	% of Primary education	-0.08	-	-0.08
	% of Secondary education	0.92	-	0.92
	% of Tertiary education	3.37	natural log (x+1)	0.69
	% of Employed	0.75	-	0.75
	% of Unemployed	0.86	-	0.86
	% of Inactive	-0.73	-	-0.73
	% of Low pay jobs	-0.71	-	-0.71
	% of Medium paid jobs	0.80	-	0.80
	% of High paid jobs	0.69	-	0.69
	% of Primary industry worker	-0.26	-	-0.26
	% of Secondary industry worker	1.03	natural log (x+1)	-0.68
	% of Tertiary industry worker	0.43	-	0.43

The socio-economic variables in Table 6.4 are considered to be important based on theoretical reviews: For example, Ravenstein's migration law indicates that female are more migratory than males; other classical migration theories such as Harris-Todaro two-sector model and Gravity model assumed migrant's decision is made based on rational economic reasons (e.g. wage, job opportunities); Dualistic Economy Model highlights the absorption of surplus labour from the agricultural sector into the industrial sector in urban areas (Young, 2004); In Britain, counterurbanisation is caused by more affluent or educated people are in search of better living conditions (Hosszú, 2009); In Germany and France, counterurbanisation was more apparent among people who aged below 18 and above 30 while other age groups displayed weak urbanisation pattern (Fielding, 1982; Kontuly & Vogelsang, 1988; Kontuly et al., 1986).

This step ends data preparation for aggregate migration modelling. The next section will explain data preparation for spatial interaction models of origin-destination flow.

6.3.2 Spatial interaction models

Datasets for this modelling approach consist of one row per census year per origin-destination District pair. In order to examine flows of one place to another, flows that involve the same origin and destination (i.e. intra-district flows) are removed. The final number of observations by census year is shown in Table 6.5.

Table 6. 5: Number of observation per origin-destination District pair.

Period	Origin District	Destination District	Number of observation
1970-80	119	127	14,994
1980-91	127	131	16,510
1991-00	131	133	17,293
Total	-	-	48,797

Table 6.5 shows the number of origin district and destination district are different within each time period. This is because of the change in geography as time passes. For example, some districts are disaggregated into two or more districts by the Malaysian government for administrative purpose. There is also an aggregation of two or more districts into one district, but the numbers that are disaggregated are greater than the number that are aggregated. As mentioned earlier, the number of observations only include the number of flows in different origin to a different destination (i.e. inter-district flows). Each origin-destination and year-specific row consists of 65 columns of variables (Table 6.6).

Table 6. 6: Variables for spatial interaction modelling

Variables	Category		Sub category
Dependant variable	Migration flows		Number of flows taken from origin to destination
Independent variables	Distance		Straight distance (km^2) from origin to destination
	Flow attributes (Dummy variables)	Flow involving changing of settlement type	Flow from core city to suburban
			Flow from core city to other area
			Flow from suburban to core city
			Flow from suburban to other suburban
			Flow from suburban to other area
			Flow from other area to core city
			Flow from other area to suburban
			Flow from other area to other area (excluded)
			Flow involving crossing sea
			Flow not involving crossing sea (excluded)
		Flow involving crossing the sea	Flow involving crossing sea
			Flow not involving crossing sea (excluded)
	Socio-economic characteristics at origin	Population	% of Total population
		Sex	% of Male
			% of Female
		Age-group	% of Child/adolescents
			% of Young adult
			% of Middle-aged adult
			% of Mature adult
			% of Elderly
			% of Working age group
		Ethnicity	% of Ethnic majority
			% of Ethnic minority
		Marital status	% of Single
			% of Married
			% of Widowed
		Education Attainment	% of Without formal education
			% of Primary education
			% of Secondary education
			% of Tertiary education
		Employment status	% of Employed
			% of Unemployed
			% of Inactive
		Types of occupation	% of Low pay jobs
			% of Medium paid jobs
			% of High paid jobs

		Types of occupation industry	% of Primary industry worker
			% of Secondary industry worker
			% of Tertiary industry worker
	Socio-economic characteristics at destination	Population	% of Total population
		Sex	% of Male
			% of Female
		Age-group	% of Child/adolescents
			% of Young adult
			% of Middle-aged adult
			% of Mature adult
			% of Elderly
			% of Working age group
		Ethnicity	% of Ethnic majority
			% of Ethnic minority
		Marital status	% of Single
			% of Married
			% of Widowed
		Education Attainment	% of Without formal education
			% of Primary education
			% of Secondary education
			% of Tertiary education
		Employment status	% of Employed
			% of Unemployed
			% of Inactive
		Types of occupation	% of Low pay jobs
			% of Medium paid jobs
			% of High paid jobs
		Types of occupation industry	% of Primary industry worker
			% of Secondary industry worker
			% of Tertiary industry worker

As previously, the data distribution needs to be checked and transform variables that are highly skewed (Table 6.7). Transformation of data is done by each period (1970-80, 1980-91 and 1991-2000) instead of the overall period (1970-2000) due to a specific reason which will be explained in Section 6.4.2.

Table 6. 7: Data transformation for spatial interaction modelling

Variables			Skewness			Best Transformation **			Skewness after transformation		
			1970-80	1980-91	1991-00	1970-80	1980-91	1991-00	1970-80	1980-91	1991-00
Dependant variable	Flow from origin to destination*		13.2	32.8	46.8	-	-	-	13.2	32.8	46.8
Independent variable	Distance (km) from origin to destination		0.2	0.3	0.3	-	-	-	0.2	0.3	0.3
	Origin	% of Total population	3.9	3.1	3.3	LN	LN	LN	0.1	0.4	0.3
		% of Male	0.6	0.8	0.5	-	-	-	0.6	0.8	0.5
		% of Female	-0.6	-0.8	-0.5	-	-	-	-0.6	-0.8	-0.5
		% of Child/adolescents	-0.4	-0.2	0.0	-	-	-	-0.4	-0.2	0.0
		% of Young adult	1.1	1.2	0.6	LN	LN	-	0.8	0.6	0.6
		% of Middle-aged adult	0.5	0.5	0.1	-	-	-	0.5	0.5	0.1
		% of Mature adult	0.3	0.2	-0.1	-	-	-	0.3	0.2	-0.1
		% of Elderly	0.4	0.2	0.1	-	-	-	0.4	0.2	0.1
		% of Working age group	0.3	0.4	0.2	-	-	-	0.3	0.4	0.2
		% of Ethnic majority	-0.5	-0.5	-0.5	-	-	-	-0.5	-0.5	-0.5
		% of Ethnic minority	0.5	0.5	0.5	-	-	-	0.5	0.5	0.5
		% of Single	-0.8	-1.2	-1.2	-	CUBE	SQ	-0.8	-0.7	-0.8
		% of Married	0.7	1.2	1.1	-	LN	LN	0.7	0.8	0.7
		% of Widowed	0.3	0.6	0.8	-	-	-	0.3	0.6	0.8
		% of Without formal education	-0.9	-0.3	0.1	-	-	-	-0.9	-0.3	0.1
		% of Primary education	0.7	0.0	-0.5	-	-	-	0.7	0.0	-0.5
		% of Secondary education	2.9	1.9	0.9	SQRT	SQRT	-	0.8	0.1	0.9
		% of Tertiary education	1.7	1.8	3.5	SQRT	SQRT	-	0.6	0.6	0.1
		% of Employed	0.5	1.0	0.8	-	LN	-	0.5	0.6	0.8
		% of Unemployed	0.8	0.8	0.7	-	-	-	0.8	0.8	0.7
		% of Inactive	-0.4	-1.0	-0.8	-	-	-	-0.4	-1.0	-0.8
		% of Low pay jobs	-1.3	-1.0	-0.7	CUBE	SQ	-	-0.5	-0.7	-0.7
		% of Medium paid jobs	1.5	1.3	0.9	LN	LN	-	0.0	0.2	0.9
		% of High paid jobs	1.0	0.4	0.5	-	-	-	1.0	0.4	0.5
		% of Primary industry worker	-1.0	-0.7	0.0	SQ	-	-	-0.2	-0.7	0.0
		% of Secondary industry worker	0.9	1.0	0.9	-	SQRT	-	0.9	0.1	0.9
		% of Tertiary industry worker	1.2	0.7	0.2	LN	-	-	-0.3	0.7	0.2
	Destination	% of Total population	3.1	3.3	3.2	LN	LN	LN	0.4	0.3	0.4
		% of Male	0.8	0.5	0.8	-	-	-	0.8	0.5	0.8
		% of Female	-0.8	-0.5	-0.8	-	-	-	-0.8	-0.5	-0.8
		% of Child/adolescents	-0.2	0.0	0.1	-	-	-	-0.2	0.0	0.1
		% of Young adult	1.2	0.6	0.8	LN	-	-	0.6	0.6	0.8
		% of Middle-aged adult	0.5	0.1	0.2	-	-	-	0.5	0.1	0.2

		% of Mature adult	0.2	-0.1	-0.5	-	-	-	0.2	-0.1	-0.5
		% of Elderly	0.2	0.1	0.5	-	-	-	0.2	0.1	0.5
		% of Working age group	0.4	0.2	0.3	-	-	-	0.4	0.2	0.3
		% of Ethnic majority	-0.5	-0.5	-0.7	-	-	-	-0.5	-0.5	-0.7
		% of Ethnic minority	0.5	0.5	0.7	-	-	-	0.5	0.5	0.7
		% of Single	-1.2	-1.2	0.1	CUBE	SQ	-	-0.7	-0.8	0.1
		% of Married	1.2	1.1	-0.2	LN	LN	-	0.8	0.7	-0.2
		% of Widowed	0.6	0.8	0.6	-	-	-	0.6	0.8	0.6
		% of Without formal education	-0.3	0.1	0.1	-	-	-	-0.3	0.1	0.1
		% of Primary education	0.0	-0.5	-0.7	-	-	-	0.0	-0.5	-0.7
		% of Secondary education	1.9	0.9	0.3	SQRT	-	-	0.1	0.9	0.3
		% of Tertiary education	1.8	3.5	2.8	SQRT	-	-	0.6	0.1	-0.3
		% of Employed	1.0	0.8	0.6	LN	-	-	0.6	0.8	0.6
		% of Unemployed	0.8	0.7	1.7	-	-	LN	0.8	0.7	0.3
		% of Inactive	-1.0	-0.8	-0.7	-	-	-	-1.0	-0.8	-0.7
		% of Low pay jobs	-1.0	-0.7	-0.3	SQ	-	-	-0.7	-0.7	-0.3
		% of Medium paid jobs	1.3	0.9	0.3	LN	-	-	0.2	0.9	0.3
		% of High paid jobs	0.4	0.5	0.8	-	-	-	0.4	0.5	0.8
		% of Primary industry worker	-0.7	0.0	0.3	-	-	-	-0.7	0.0	0.3
		% of Secondary industry worker	1.0	0.9	0.7	SQRT	-	-	0.1	0.9	0.7
		% of Tertiary industry worker	0.7	0.2	0.1	-	-	-	0.7	0.2	0.1

Note:
** Flow is not transformed although it is highly skewed in order to perform non-linear regressions (Poisson and Negative binomial) that require counts variable.*
*** LN = natural log*
SQRT = square root
SQ = square
CUBE = cube

This step ends data preparation for Spatial Interaction Modelling. The next section will discuss methods on how to perform the modelling starting with aggregate migration flow models in Section 6.4.1 followed by origin-destination flow models in Section 6.4.2.

6.4 Methods

6.4.1 Aggregate flow models

The modelling process uses three types of regression methods: Ordinary least square, Poisson and finally Negative binomial regressions. Three different regression models are used to test which approach is better in explaining the aggregate migration flow. For example, Ordinary least square regression assumes the dependent variable (e.g. in-migration) has a normal (Gaussian) distribution. However, since the values of in-migration and out-migration cannot be negative, the distribution normally is asymmetrical and highly skewed with a long tail to the right, hence resembling Poisson distribution. Negative binomial on the other hand has similar distribution as Poisson and both models used the same type of dependant variable: Both models deal with count dependant variable but Negative binomial model usually is used when the outcome of Poisson model is over-dispersed. Over-dispersion happens when the conditional variance exceeds the conditional mean (UCLA: Statistical Consulting Group, n.d.) or residual deviance of a model exceeds degrees of freedom (Dormann, 2016). This analysis can be fitted as follows:

$$\textit{Net migration}^{t\text{-}t+1} \sim \textit{year} + \textit{settlement\_type} + \textit{total\_pop}^{\,t} + \textit{socioeconomic}^{\,t}$$

$$\textit{In-migration}^{\,t\text{-}t+1} \sim \textit{year} + \textit{settlement\_type} + \textit{total\_pop}^{\,t} + \textit{socioeconomic}^{\,t}$$

$$\textit{Out-migration}^{\,t\text{-}t+1} \sim \textit{year} + \textit{settlement\_type} + \textit{total\_pop}^{\,t} + \textit{socioeconomic}^{\,t}$$

Where *Net migration*$^{t\text{-}t+1}$ and *In-migration* $^{t\text{-}t+1}$ and *Out-migration* $^{t\text{-}t+1}$ are the net migration, in-migration and out-migration that occurs from one time period to the next, *year* is dummy variables for census year, *settlement_type* is dummy variables for types of settlement, *total_pop*^{t} is the total population at the starting year, and *socioeconomic*$^{\,t}$ are the best sets of socioeconomic attributes at the starting year.

In order to obtain the best set of socioeconomic attributes, this chapter refers to the guideline provided by Field (2009). Backward regression method is used to retain the most important variables and remove those that are less or not significant. This method starts by inserting all explanatory variables into the model and it calculates the contribution of each variable based on their significance value. If a variable is not statistically significant, then it will be removed from the model and the remaining variables are re-assessed. Besides the backward method, there is also a forward method. The forward method is known to operate opposite to the backward method. The forward method starts with the null model and selects the most significant variable first and then look for the second best, third and so forth. However, the backward method is preferable rather than the forward method because of the suppressor effects. The suppressor effects are more likely happens when using the forward method when the 'predictor has a significant effect but only when another variable is held constant' (Field, 2009). Therefore, the forward method is prone to a higher risk to not include a significant variable into the model.

However, there are also several shortcomings in both methods. The methods are arguably to not consider the researcher's decisions on which variables should best predict the outcome. This situation may lead to the differences between the outcome and theoretical findings. Furthermore, there is a possibility that the methods may capture too many variables which have a small contribution to the outcome (over-fitting) or may not include important variables into the model (under-fitting).

The main reason the backward method is chosen is that all explanatory variables selected in Section 6.3 are known to have theoretical importance. For example, Ravenstein's migration law indicates that female are more migratory than males and migration is due to more commerce activities in towns; neo-classical migration theory such as Harris-Todaro two-sector model assumed migrant's decision is made based on rational economic reasons (e.g. wage, job opportunities); and the dualistic economy model highlights the absorption of surplus labour from the agricultural sector into the industrial sector in urban areas (Young, 2004).

The quality of the model will be examined based on the model fit: R-Squared and *Akaike Information Criterion*: The higher the R-squared, the better because more variation between the dependant and independent variables are explained in the model; On the contrary, the *Akaike Information Criterion* also known as AIC is mainly used to asses quality of non-linear model (Poisson and Negative binomial) where lower the AIC the better. However, the AIC of a model can only be compared to a model that is similar type (e.g. the AIC of an Ordinary least square model cannot be compared to AIC of a Poisson model or Negative binomial model). Furthermore, the AIC can only be compared to a model that has a similar dependant variable (e.g. In-migration to In-migration).

Besides observing R-squared and AIC, another alternative is to check the quality of the model is by performing predictive checking test. According to Andrew & Jennifer (2006), the purpose of predictive checking test is to assess errors or uncertainty produced in the model by simulating several scenarios. The function takes a model and the set of independent variables used and return predictions from the model. The outcome of the predictive checking test will show a comparison between the distribution of actual values of dependant variable and the expected values as well as the simulated values. If the simulated values or curve is close to the expected values, the proposed model is deemed to better fit the underlying data. In contrast, if the simulation values or the curve is a mismatch to the expected values, then the model is deemed to be poor to the underlying data

6.4.2 Spatial interaction models

There are several steps taken to build to a model of the flow between origin and destination: the process starts with a) unconstrained null model; followed by b) doubly-constrained model; c) origin-constrained model; d) destination-constrained model; and finally developed into an e) unconstrained model. A detailed explanation of these models will be explained in the respective topics. As noted in the previous section, the quality of each model created will be assessed based on the values of R-squared for Ordinary least square model, AIC for Poisson and Negative binomial models, and predictive checking analysis.

a) Unconstrained null model

Before modelling spatial interaction, it is important to understand the basic relationship between flow and distance using an unconstrained null model. The reason it is called 'unconstrained null model' is that it places no constraint on predicted flows, in fact, flow is solely explained by the distance between origin and destination. From a theoretical perspective, flow is expected to decay as distance increases. The linear relationship can be fitted using a simple Ordinary least square regression:

$$Flow_{ij} \sim d_{ij}$$

Where $Flow_{ij}$ is flows taken from origin to destination and d_{ij} is the straight distance between the origin-destination pair involved in the flow. Once the basic relationship between flow and distance is understood, the next step is to constrain the model by incorporating origin and destination attributes.

b) Doubly constrained model

A traditional doubly constrained model includes dummy variables for each origin and for each destination. It is not possible to include any other origin or destination attributes (such as total population at origin or destination) since the effect of all of these is captured by the single 'origin' and 'destination' dummy variables (regression coefficient). The model can be fitted as follows:

$$Flow_{ij} \sim origin + destination$$

Where *origin* and *destination* represent dummy variables for the origin and destination Districts.

The above provides a 'null' model against which to test more sophisticated models. It is known as 'null' model because it will correctly predict the total flows leaving each origin and arriving at each destination with 100 percent accuracy (Hence ‘doubly constrained’). However, the model will not predict the size of the flows between each origin-destination pair with 100 percent accuracy (because the size of individual origin-destination flows is not constrained). Figure 6.1 shows the matrix representation of flows for the doubly constrained model.

District		destination$_j$				predicted total Flow$_i$ constrained to origin$_i$
		1	2	3	4	
origin$_i$	1	Flow$_{ij}$ ~ origin + destination				Flow$_{i1}$
	2					Flow$_{i2}$
	3					Flow$_{i3}$
	4					Flow$_{i4}$

predicted total Flow$_j$ constrained to destination$_j$	Flow$_{j1}$	Flow$_{j2}$	Flow$_{j3}$	Flow$_{j4}$	Total

Figure 6. 1: Matrix representation of flows for the doubly constrained model

Because all origin and destination attributes have already been controlled for in a doubly constrained model, the purpose of such a model is clearly not to explore the influence of origin and destination attributes on the size of flows. Rather, the purpose of a doubly constrained model is to explore how attributes of the flow itself influence the size of the flow – e.g. the length of the flow (distance between origin and destination).

c) Selecting the best regression method

For the same reason as aggregate migration modelling, a doubly constrained model should be fitted separately using each of Ordinary least square, Poisson or Negative binomial models. Normally, flows cannot be negative and the distribution is asymmetrical, opposite to the Ordinary least square regression assumption of a normal distribution. Hence, this requires further observation on Poisson regression since it has a similar pattern of distribution. However, if the outcome of the Poisson model is over-dispersed, one solution is to perform the

Negative binomial regression since it has an extra parameter to model over-dispersion.

The key attribute associated with each flow is the length (distance) of the flow. Another might be a dummy variable indicating whether or not the flow 'involves crossing a sea'. These flow attributes can be added to the model, to explore the nature of their influence and their relative importance. For now, this section will focus on exploring distance decay. This can be explored by simply adding a distance decay term to the model:

$$Flow_{ij} \sim origin + destination + d_{ij}$$

The signs of the distance decay (d_{ij}) coefficient indicates the nature of its influence on the size of flows. The fit of the above models can be measured by observing the R-squared, AIC and simulation test. This will allow an assessment of the relative fit of each model, and hence identification of the best combination of distributional assumptions and measure of distance.

d) Selecting the best way to explain years of flow

There is, however, one problem with the previous model. It assumes that there is only one observed flow per origin-destination pair while there are four flows per origin-destination pair (one per Census Year). Hence the models outlined above won't be fully doubly constrained. The next step is to identify the best way to explain years of flow.

Two solutions are proposed for this problem. First, the models are fitted separately for each Census year. The equation is as follows:

$$Flow_{ijt} \sim origin + destination + d_{ij}$$

Where t is a census year. The purpose of this approach is that it can be shown how the coefficients associated with the flow attributes (e.g. rate of distance decay) have changed (or not) over successive censuses. However, the results will be less robust (wider confidence intervals around model coefficients) due to the reduction in the size of the dataset through the exclusion of flows from other years.

Second, all years are included in the model at the same time by including dummy variables for Year which can be fitted as follows:

$$Flow_{ij} \sim origin + destination + d_{ij} + year$$

Where *year* are dummy variables for years of flow (*year8091* and *year9100* excluding *year7080).* This model assumes *origin*, *destination* and d_{ij} are constant throughout the years (1970-2000), and hence it is different than the previous model.

The purpose of the second solution is to find out how well this approach can capture the year-specific effects identified via the first solution. If it captures the effects reasonably well, the second approach is preferable because it simplifies the reporting of results (only one model covering all years instead of one per census year). The second approach should also lead to narrower confidence intervals (more certain results) because the results are based upon a larger sample size (all flows instead of just the flows for a given census year).

e) Selecting more flow attributes

Once have identified the best way to explain years of flow, the next set up is to improve the model by incorporating more flow attributes that influence the size of flows. The attributes that are examined in the model are *crossing_sea* (a dummy variable indicating whether the flow involves crossing the sea or not) and *core_sub, core_other, sub_sub, sub_core, sub_other, other_core* and *other_sub* (dummy variables recording the change of settlement type involved in the flow).

These attributes are added because Malaysia geographically is divided into the South and West regions. Movement between these regions requires crossing the sea. Furthermore, types of settlement plays an important role to the migration process in Malaysia based on results in Chapter 5 on internal migration where *core_sub* is flow from core city to suburban, *core_other is* flow from core city to other area, *sub_sub* is flow suburban to other suburban, *sub_core* is flow from suburban to core city, *sub_other* is flow from suburban to other area, *other_core* is flow from other area to core city and *other_sub* is flow from other area to suburban.

In order to examine these attributes thoroughly, seven model variants are built as shown in Table 6.8:

Table 6. 8: Seven model variants to select flow attributes

No.	Models	Explanation
1	*Flowij ~ origin + destination + crossing_sea*	To check if there is a 'crossing the sea' effect.
2	*Flowij ~ origin + destination + core_sub + core_other + sub_sub + sub_core + sub_other + other_core + other_sub*	To check if there is 'change of settlement type' effect.
3	*Flowij ~ origin + destination + dij + crossing_sea*	To check if there is 'crossing the sea' effect persists once distance of flow has been taken into account.
4	*Flowij ~ origin + destination + dij + core_sub + core_other + sub_sub + sub_core + sub_other + other_core + other_sub*	To check if there is 'change of settlement type' effect persists once distance of flow has been taken into account.
5	*Flowij ~ origin + destination + dij + crossing_sea + core_sub + core_other + sub_sub + sub_core + sub_other + other_core + other_sub*	To check whether having taken account of flows distance, and a 'crossing the sea' effect, a 'change of settlement type' effect persists.
6	*Flowij ~ origin + destination + dij + crossing_sea + dij*crossing_sea*	To check for possible interactions between the various flows attributes.
7	*Flowij ~ origin + destination + dij + core_sub + core_other + sub_sub + sub_core + sub_other + other_core + other_sub +*	

	*dij*core_sub + dij*core_other + dij*sub_sub + dij*sub_core + dij*sub_other + dij*other_core + dij*other_sub*	

Once have identified the best way in explaining census years and best sets of flow attributes, the next step is to observe push and pull factors of migration flow.

f) Origin-constrained model

An origin-constrained model correctly predicts the total number of flows from a given origin. However, it makes no constraint on the number of flows to each predicted destination. Instead, destination attributes that might explain the size of the flows to each destination are added to the model. The model can be fitted as follows:

$$Flow_{ij} \sim origin + d_{ij} + core\_sub + core\_other + sub\_sub + sub\_core + sub\_other + other\_core + other\_sub + crossing\_sea + socioeconomic_j$$

Where $socioeconomic_j$ are a set of socioeconomic attributes at the destination. The purpose of this model is to find the best sets of socioeconomic attributes that attract people to move to a destination. Figure 6.2 shows the matrix representation of flows for the origin-constrained model.

<table>
<tr><td colspan="2" rowspan="2">District</td><td colspan="4">destination$_j$</td><td></td><td rowspan="2">predicted total Flow$_i$ constrained to origin$_i$</td></tr>
<tr><td>1</td><td>2</td><td>3</td><td>4</td><td></td></tr>
<tr><td rowspan="4">origin$_i$</td><td>1</td><td colspan="4" rowspan="4">Flow$_{ij}$ ~ origin… socioeconomic$_j$</td><td></td><td>Flow$_{i1}$</td></tr>
<tr><td>2</td><td></td><td>Flow$_{i2}$</td></tr>
<tr><td>3</td><td></td><td>Flow$_{i3}$</td></tr>
<tr><td>4</td><td></td><td>Flow$_{i4}$</td></tr>
<tr><td colspan="2">unconstrained predicted total Flow$_j$</td><td>Flow$_{j1}$</td><td>Flow$_{j2}$</td><td>Flow$_{j3}$</td><td>Flow$_{j4}$</td><td></td><td>Total</td></tr>
</table>

Figure 6. 2: Matrix representation of flows for the origin-constrained model

g) Destination-constrained model

In contrast to the origin-constrained model, destination-constrained model correctly predicts the total number of flows to a given destination and makes no constraint on the number of flows from each predicted origin. Instead, origin attributes that might explain the size of the flow to each origin is considered. The model can be fitted as follows:

$$Flow_{ij} \sim destination + d_{ij} + core\_sub + core\_other + sub\_sub + sub\_core + sub\_other + other\_core + other\_sub + crossing\_sea + socioeconomic_i$$

Where $socioeconomic_i$ are the combination of socioeconomic attributes at the origin. The purpose of this model is to find the best sets of socioeconomic attributes that pushes people to move out from an origin. Figure 6.3 shows the matrix representation of flows for the destination-constrained model.

<table>
<tr><td colspan="2" rowspan="2">District</td><td colspan="4">destination_j</td><td></td><td rowspan="2">unconstrained predicted total Flow_i</td></tr>
<tr><td>1</td><td>2</td><td>3</td><td>4</td><td></td></tr>
<tr><td rowspan="4">origin_i</td><td>1</td><td colspan="4" rowspan="4">Flow_{ij} ~ destination… socioeconomic_i</td><td></td><td>Flow_{i1}</td></tr>
<tr><td>2</td><td></td><td>Flow_{i2}</td></tr>
<tr><td>3</td><td></td><td>Flow_{i3}</td></tr>
<tr><td>4</td><td></td><td>Flow_{i4}</td></tr>
<tr><td colspan="2">predicted total Flow_j constrained to destination_j</td><td>Flow_{j1}</td><td>Flow_{j2}</td><td>Flow_{j3}</td><td>Flow_{j4}</td><td></td><td>Total</td></tr>
</table>

Figure 6. 3: Matrix representation of flows for the destination-constrained model

h) Unconstrained model

An unconstrained model is the final spatial interaction model considered. It places no constraints on the predicted size of the flows between each origin and destination. The only constraint it imposes is that the total predicted flows between all origins and destinations equals the observed total flows between all origins and destinations. Instead, the predicted flows in the model between each origin-destination pair are driven by the attributes added to the model describing origins, destinations and flows. This model can be fitted as follows:

$$Flow_{ij} \sim d_{ij} + core\_sub + core\_other + sub\_sub + sub\_core + sub\_other + other\_core + other\_sub + crossing\_sea + socioeconomic_i + socioeconomic_j$$

Where $socioeconomic_i$ and $socioeconomic_j$ are the best sets of origin and destination socio-economic attributes obtained from the origin-constrained and destination-constrained models. Figure 6.4 shows the matrix representation of flows for the unconstrained model.

District		destination$_j$ 1	2	3	4	unconstrained predicted total Flow$_i$
origin$_i$	1	Flow$_{ij}$ ~ … socioeconomic$_i$ + socioeconomic$_j$				Flow$_{i1}$
	2					Flow$_{i2}$
	3					Flow$_{i3}$
	4					Flow$_{i4}$
unconstrained predicted total Flow$_j$		Flow$_{j1}$	Flow$_{j2}$	Flow$_{j3}$	Flow$_{j4}$	Total

Figure 6. 4: Matrix representation of flows for the unconstrained model

i) The spatial pattern of model residual

Finally, this chapter will map out the spatial pattern of the model residuals by origin and destination. The process is as follow: 1) Find model residuals, 2) Sum by origin, 3) Sum by destination, 4) Map model residuals by origin and 5) Map model residuals by destination. In order to check whether the residuals show a specific spatial pattern or not, a spatial autocorrelation test is used. Spatial autocorrelation is built upon the first law of geography (*"Everything is related to everything else, but near things are more related than distant things."*) and can be measured by calculating Moran's I (Index) (Tobler, 1970). The ArcGIS software allows us to automatically calculate spatial autocorrelation of each residual maps (by origin and destination). The result can be classified either positive, negative or no spatial autocorrelation: positive spatial autocorrelation is when the index is near to +1 and similar values cluster together in a map; negative spatial autocorrelation is when the index is near to -1 and dissimilar values cluster together in a map; no spatial correlation is when the index is 0 and the values are randomly distributed.

6.4.3 Software and dependencies

The primary software that is used for modelling is RStudio (RStudio Team, 2016). Besides, Statistical Package for the Social Science (SPSS) (Version 24.0, 2016) software is also used to validate results produced by RStudio. Several dependencies need to be installed in RStudio first before modelling:

1. '*foreign*'(R package version 0.8-67, 2016) – Since the dataset is in SPSS format, this package is required in order to read the dataset in the RStudio
2. '*lme4*' (Bates, Maechler, Bolker, & Walker, 2015) – to perform Ordinary least square regression and Poisson regression
3. '*MASS*' (Venables & Ripley, 2002) – to perform Negative binomial regression
4. '*nlme*' (R package version 3.1-128, 2016) – to check significance of explanatory variables in non-linear model
5. '*car*'(J. Fox & Weisberg, 2011) – to check multicollinearity
6. '*arm*' (R package version 1.10-1, 2018) – to do simulation for predictive checking
7. 'ggplot2'(Wickham, 2016) – to plot graphs

A spatial autocorrelation test was implemented using ArcGIS within the Spatial Statistics Tools through Spatial Autocorrelation Moran's I (Index).

6.5 Aggregate flows models

This section discusses the results that arise from models of migration flows at the aggregate level (net migration, in-migration, and out-migration). However, before presenting these results, it is important to understand the observed pattern of migration flows in Malaysia. Figure 6.1 depicts a density plot of net migration, in-migration, and out-migration for each year during 1980-2010.

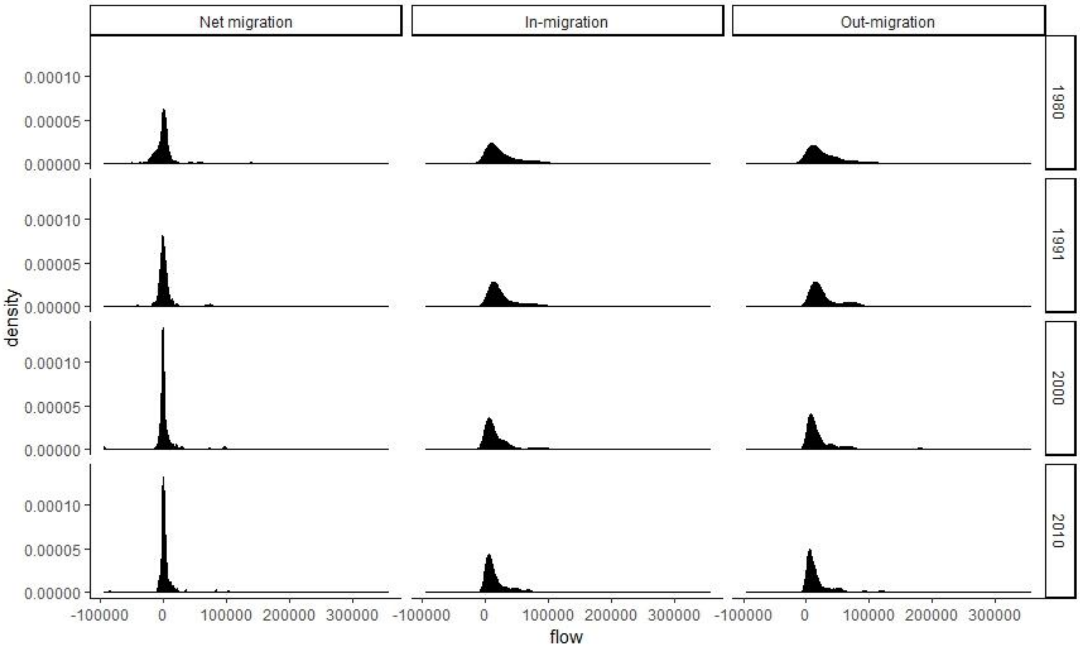


Figure 6. 5: Density plot of net, in-, and out-migration, 1980-2010

Figure 6.5 shows that the distribution of flow sizes for all migration types are positively skewed, indicating a non-normal data distribution. As mentioned in Section 6.4.1., to perform ordinary least squares regression, the migration values need to be transformed because this type of regression assumes the values are normally distributed. For Poisson and negative binomial regression models, there is no need for transformation because these models deal with count data. However, since net migration contains negative values, these must be transformed into positive values to perform Poisson and negative binomial models by adding an equal amount of positive values to the negative values. For example, if the lowest net migration value is -100, it can be transformed into a positive value by adding 101 to all net migration values. This transformation leaves the shape of the distribution, and its skewness, unchanged.

The following sections discuss the results for net migration models (Section 6.5.1), in-migration models (Section 6.5.2), and out-migration models (Section 6.5.3).

6.5.1 Net migration model

Table 6. 9: Net migration models

Variables/ Models	Model 1 (Ordinary least square)		Model 2 (Poisson)		Model 3 (Negative binomial)	
	B	Sig	B	Sig	B	Sig
(Intercept)	295.943	0.000	10.950	0.000	11.166	0.000
Years (dummy)						
1970-1980	*	*	*	*	*	*
1980-1991	1.292	0.603	0.000	0.842	0.006	0.822
1991-2000	-2.093	0.436	-0.020	0.000	0.003	0.902
2000-2010	-2.663	0.404	-0.052	0.000	0.016	0.530
Settlement type (dummy)						
Capital metropolitan core	-214.983	0.000	-1.666	0.000	-1.627	0.000
Capital metropolitan suburban	38.071	0.000	0.250	0.000	0.252	0.000
Regional metropolitan	14.050	0.001	0.096	0.000	0.084	0.036
Intermediate cities	-6.908	0.044	-0.049	0.000	-0.035	0.301
Small towns/ villages	-6.019	0.012	-0.030	0.000	-0.039	0.098
Remote villages	*	*	*	*	*	*
Attributes at starting year						
Total population (log)	1.452	0.299	0.017	0.000	0.013	0.337
% Male			0.004	0.000		
% Young adult			0.001	0.000	0.006	0.017
% Middle-aged adult			0.003	0.000		
% Mature adult	-0.898	0.021	-0.005	0.000		
% Ethnic majority			0.001	0.000		
% Married			0.002	0.000		
% Attained primary education			0.000	0.000		
% Attained secondary education			0.010	0.000		
% Attained tertiary education ((log)x+1)	8.348	0.000	0.037	0.000		
% Employed			0.000	0.003		
% Unemployed			-0.007	0.000		
% Medium-paid workers			0.001	0.000		
% Secondary industry workers ((log)x+1)			-0.012	0.000		

R-squared	0.534		
AIC		687,008	11,485
Residual deviance		680,205	486
Degrees if freedom		559	498

Note:
** The variable was removed to avoid multicollinearity*

1. *Each model captures different socio-economic attributes (e.g., the ordinary least squares model captures two socio-economic attributes)*
2. *Each model uses different dependent variables (net migration) due to data transformation and regression requirements (Poisson and negative binomial regressions require count and positive values of dependent variables for modelling):*
   - *Ordinary least squares: square root (net migration + 94,248)*
   - *Poisson: net migration + 94,248*
   - *Negative binomial: net migration + 94,248*

Table 6.9 shows three variations of net migration models using the ordinary least squares, Poisson, and negative binomial regressions. While year and settlement types are fixed in all models, different models retain different socio-economic attributes. Notably, the Poisson model, after backward elimination, retains the most attributes. Despite this difference in the models, the pattern of coefficients is similar for attributes appearing in multiple models (e.g., the percentage of mature adults has a negative coefficient in the ordinary least squares and Poisson models, and the percentage of young adults has a positive coefficient in the negative binomial and Poisson models).

To determine which model is better at explaining net migration, a predictive checking test was undertaken for the ordinary least squares model and the Poisson model. The negative binomial was not included in the test because the R package used to implement predictive checking does not work for the negative binomial model. However, if the Poisson model captures net migration well, then there is no need to perform negative binomial regression because Poisson regression is sufficient for measuring count data.

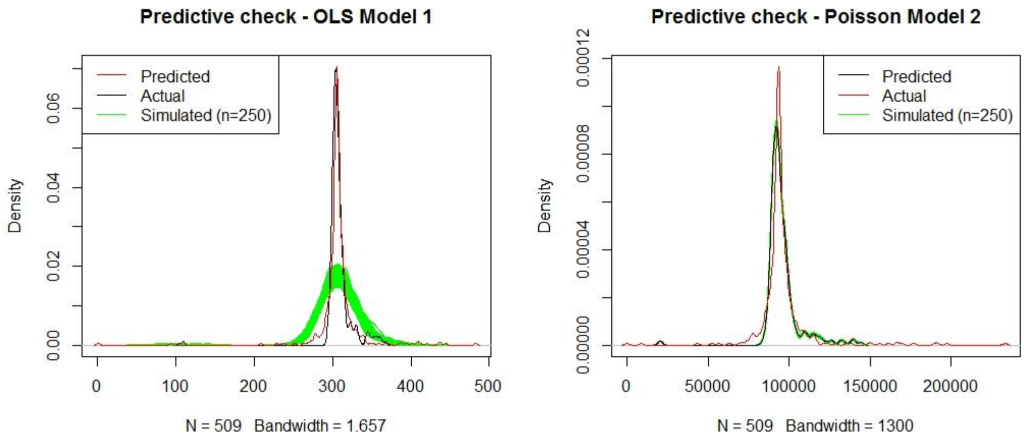


Figure 6. 6: Predictive check simulation of net migration models

Figure 6.6 shows the simulated values or curves generated from random predictions of the model. The simulated curve for the ordinary least squares model clearly does not fit the actual or predicted values, over-smoothing the peak in the distribution. This is because the poor model fit leads to a large error or uncertainty caused by a large amount of unexplained variation in the model (Andrew & Jennifer, 2006). On the contrary, the simulated curve for the Poisson model fits the predicted values well. Based on the predictive checking test, the Poisson model clearly is better than ordinary least squares model.

However, there is one major problem with the Poisson model. Poisson regression assumes that variance is equal to mean (ratio of 1:1), and thus residual deviance should be equal to residual degrees of freedom. The difference between residuals and degrees of freedom are very large in the Poisson model (680,204 to 559), indicating an over-dispersed distribution. Over-dispersion can be addressed by assuming a negative binomial rather than a Poisson distribution.

The result of the negative binomial model shows a similar ratio of residual deviance and degrees of freedom (559 to 498). This means the negative binomial model is better at predicting net migration and at addressing over-dispersion than the Poisson model. However, despite its goodness of fit, only one socio-economic attribute is associated with net migration (the percentage of young adults). This may be because the model struggles to capture net migration flow, which is the combined outcome of the two competing phenomenon: in-migration and out-migration. As Rogers (1990) observed, there is no such thing as a net migrant.

There are three significant variables based on the negative binomial results in Table 6.9: capital metropolitan core city, capital metropolitan suburban areas, and area with young adults. The coefficients for the capital metropolitan core city and suburban areas show a contrasting pattern: the core city has a negative coefficient while the suburban areas has a positive coefficient. The result relates to the findings in the previous chapter on internal migration that capital metropolitan suburban areas had the largest net-inflows. In contrast, there were large net-outflows in the capital metropolitan core. The dominance of *Kuala Lumpur* as the core city in Malaysia has deteriorated since 1980 as a result of rapid suburbanisation in the adjacent areas. Many new townships, multinational companies, residential areas, commercial and industrial centres, and major infrastructure were built in the suburban areas. This also resulted from the overflow of urban development from *Kuala Lumpur* since the city is fully urbanised and confined by limited space.

Another important finding is the positive coefficient of the percentage of young adults. Unlike other age groups, rural-urban movement was notably evidenced among young adults originating from small towns/villages since 1980. The migration of young adults to larger settlements (particularly to the capital metropolitan area) was generally driven by economic motivations such as wage and job opportunities. Large cities such as Kuala Lumpur and its suburban areas are the most urbanised areas in Malaysia and hence provide vast economic opportunities. However, the net migration model is not able to capture additional economic attributes such as types of occupation and industries in the observed units.

The next section discusses in-migration (Section 6.5.2) and out-migration (Section 6.5.3).

6.5.2 In-migration model

Table 6. 10: In-migration models

Variables/ Models	Model 1 (Ordinary least square)		Model 2 (Poisson)		Model 3 (Negative binomial)	
	B	Sig	B	Sig	B	Sig
(Intercept)	-5.433	0.000	-2.348	0.000	-1.769	0.000
Years (dummy)						
1970-1980	*	*	*	*	*	*
1980-1991	-0.257	0.000	-0.267	0.000	-0.155	0.000
1991-2000	-1.129	0.000	-0.935	0.000	-0.981	0.000
2000-2010	-1.350	0.000	-1.433	0.000	-1.523	0.000
Settlement type (dummy)						
Capital metropolitan core	0.062	0.785	-0.096	0.000	-0.211	0.322
Capital metropolitan suburban	0.215	0.016	0.226	0.000	0.210	0.011
Regional metropolitan	0.211	0.009	0.224	0.000	0.160	0.028
Intermediate cities	0.126	0.061	0.137	0.000	0.027	0.683
Small towns/ villages	-0.030	0.527	0.021	0.000	-0.035	0.415
Remote villages	*	*	*	*	*	*
Attributes at starting year						
Total population (log)	1.039	0.000	1.017	0.000	0.908	0.000
% Male	0.037	0.000	0.028	0.000	0.031	0.000
% Child/ adolescents	0.016	0.010				
% Young adult			-0.004	0.000		
% Elderly			0.005	0.000		
% Ethnic majority			0.001	0.000		
% Married	0.023	0.001	0.013	0.000		
% Widowed/ Separated			-0.094	0.000		
% Attained primary education	0.017	0.000	-0.002	0.000		
% Attained secondary education			-0.002	0.001		
% Unemployed	-0.036	0.000	-0.068	0.000	-0.063	0.000
% Inactive			-0.005	0.000		
% Medium-paid workers					0.022	0.000
% High-paid workers			0.002	0.000		
% Secondary industry workers ((log)x+1)			-0.024	0.000		
% Tertiary industry workers			0.005	0.000		
R-squared		0.904				
AIC				1,147,722		10,132
Residual deviance				1,141,859		518
Degrees if freedom				486		496

Note:
** The variable was removed to avoid multicollinearity*

1. *Each model captures different socio-economic attributes (e.g., the ordinary least squares model captures five socio-economic attributes)*
2. *Each model uses different dependent variables (in-migration) due to data transformation and regression requirements (Poisson and negative binomial regressions require count and positive values of dependent variables for modelling):*
    - *Ordinary least squares: natural log (in-migration)*
    - *Poisson: actual values of in-migration*
    - *Negative binomial: actual values of in-migration*

Based on Table 6.10, the number of explanatory variables of each in-migration model is similar to the net migration model: the Poisson model captures a larger number of socio-economic attributes than the ordinary least squares and negative binomial models. Furthermore, the coefficients for a given attribute show a similar pattern among models (e.g., the coefficient of the male population shows positive values for all models; similarly, the coefficient of the unemployed population shows negative values for all models). The only difference spotted between the models is the opposite coefficient of people who attained primary education (i.e., positive in ordinary least squares model and negative in the Poisson model).

The R-squared value of the ordinary least squares model shows an almost perfect model fit: 90 percent of the variation of in-migration is explained by the model compared to only 53 percent for the net migration model. Again, predictive checking was done to check the quality of the models.

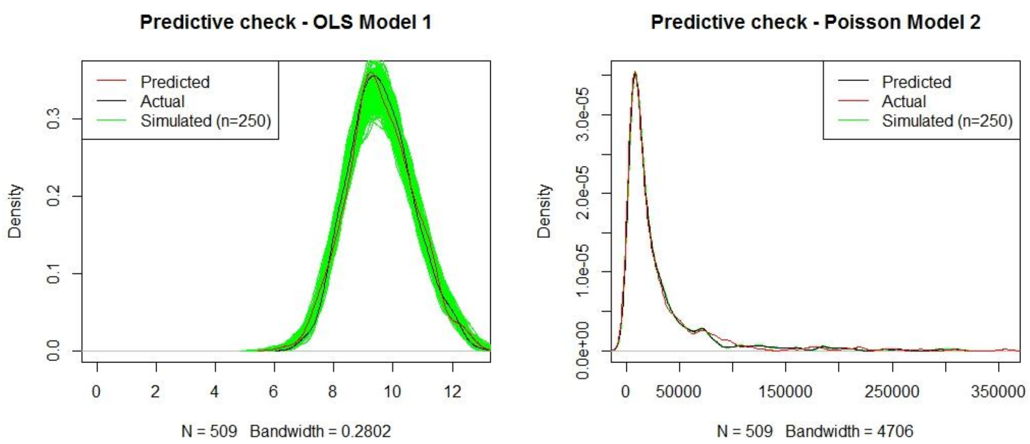


Figure 6. 7: Predictive check simulation of in-migration models

Figure 6.7 shows the predictive checking test between the ordinary least squares model and Poisson models. The Poisson model captures the simulation curve better than the ordinary least squares model although both predict the actual values reasonably well. As mentioned in the previous section, although the Poisson model is considered better than the ordinary least squares regression model, there is a large gap between residual values and degrees of freedom (1,141,859 to 486), again indicating an over-dispersion. Thus, negative binomial modelling is required to tackle this issue.

The results for the negative binomial model in Table 6.2 show the value residual deviance is close to the degrees of freedom (518 to 496), meaning over-dispersion is handled appropriately. There are several significant explanatory variables in the model: all years, capital metropolitan suburban areas, percentage of the male population, percentage of unemployed, and percentage of medium-paid workers. The increasing negative coefficient for years means that in-migration is expected to decline as the year increases. This relates to the findings in Chapter 5 (Section 5.3) whereby the values of in-migration gradually decline as year increases for all settlement types. Furthermore, the positive coefficient of in-migration in capital metropolitan suburban areas is because this settlement type is known to have received the highest amount of in-migration from lower-level settlements; suburban areas had a major increase of in-migration from 330,000 to 410,000 of migrants during 1980-2010. As explained in the previous section and chapters, this was primarily due to the rapid suburbanisation process in the capital metropolitan city.

In terms of socio-economic attributes, the positive coefficient of the percentage of male population and medium-paid workers indicate that more in-migration is expected to happen if there are more males and medium-paid workers in a district. Medium-paid occupations are jobs that provide service to the customer such as technicians, associate professionals, clerks, service workers and shop and market sales. These jobs commonly are found everywhere either in rural or urban areas in Malaysia. According to Del Carpio, Özden, Testaverde, and Wagner (2015), in-migration is primarily caused by internal movement among natives, especially those who worked in middle and lower-skilled occupations. The negative coefficient of the percentage of unemployed people on the other hand indicates that in-migration is expected to be less in areas that have a high unemployment rate. This is a common situation whereby people will migrate to areas that have high employment (typically in cities) rather than to area that has low employment or high unemployment. One of the theories that might relate to this situation is the two-sector model by Harris and Todaro (1970). The model assumes rural migrants tend to be attracted with the expectation of higher earnings and willing to accept lower wages and unemployed risk in urban areas. Also, the migrants will still prefer to migrate if they think there is a possibility or potential of getting more income in the future (Mihermutu, 2011).

These models however are not sufficient to capture the full determinants of in-migration. This is probably because there might be other unexplained factors or reasons or insufficient explanatory variables to explain why in-migration is happened. For example, based on studies done by the UNICEF, following family is known to be the primary reason of internal migration in Malaysia followed by environmental aspect, career, marital status, and education status (UNESCO et al., 2012). The ordinary least squares model and Poisson model did capture marital and education status as significant determinants of in-migration, but since they are not statistically correct, both models need to be rejected. This might be because of the different approaches used (qualitative versus quantitative).

6.5.3 Out-migration model

Table 6. 11: Out-migration models

Variables/ Models	Model 1 (Ordinary least square)		Model 2 (Poisson)		Model 3 (Negative binomial)	
	B	Sig	B	Sig	B	Sig
(Intercept)	-2.271	0.000	-1.232	0.000	-1.603	0.000
Years (dummy)						
1970-1980	*	*	*	*	*	*
1980-1991	-0.140	0.000	-0.277	0.000	-0.144	0.000
1991-2000	-0.981	0.000	-0.923	0.000	-0.973	0.000
2000-2010	-1.664	0.000	-1.392	0.000	-1.511	0.000
Settlement type (dummy)						
Capital metropolitan core	-0.027	0.880	0.015	0.000	0.087	0.631
Capital metropolitan suburban	-0.296	0.000	-0.267	0.000	-0.280	0.000
Regional metropolitan	-0.033	0.611	-0.070	0.000	0.007	0.916
Intermediate cities	-0.053	0.347	-0.022	0.000	0.034	0.545
Small towns/ villages	0.015	0.681	0.034	0.000	0.029	0.433
Remote villages	*	*	*	*	*	*
Attributes at starting year						
Total population (log)	1.010	0.000	1.061	0.000	1.004	0.000
% Male			0.001	0.000		
% Child/ adolescents	0.015	0.000			0.009	0.014
% Young adult			-0.008	0.000		
% Mature adult	0.027	0.000	0.008	0.000	0.013	0.038
% Ethnic majority			0.000	0.000	-0.002	0.057
% Married			-0.001	0.001		
% Widowed/ Separated			-0.030	0.000		
% Attained primary education			-0.001	0.000		
% Attained secondary education			-0.001	0.018		
% Attained tertiary education ((log)x+1)	0.210	0.000			0.256	0.000
% Unemployed	-0.032	0.000	-0.030	0.000	-0.031	0.000
% Inactive			-0.005	0.000		
% Medium-paid workers	0.016	0.000	0.010	0.000		
% High-paid workers			0.008	0.000		
% Secondary industry workers ((log)x+1)	-0.051	0.035	-0.079	0.000	-0.062	0.010
% Tertiary industry workers					0.004	0.013
R-squared		0.937				
AIC				690,204		9,958
Residual deviance				684,340		516
Degrees if freedom				486		492

Note:

* *The variable is removed to avoid multicollinearity*

1. *Each model capture different socio-economic attributes (e.g., ordinary least squares model only capture six socio-economic attributes).*
2. *Each model uses different dependant variable (out-migration) due to data transformation and regression type requirement (Poisson and Negative binomial regressions require count and positive values of dependant variable for modelling):*
    - *Ordinary least square: natural log (out-migration)*
    - *Poisson: actual values of out-migration*
    - *Negative binomial: actual values of out-migration*

Based on Table 6.11, the out-migration model produced a similar structure of explanatory variables as net migration and in-migration model: the Poisson model captures more socio-economic attributes than the ordinary least squares and negative binomial models and the coefficient values between the three models are similar (e.g., positive coefficient for percentage of mature adult and negative coefficient for percentage of unemployed people and secondary industry workers).

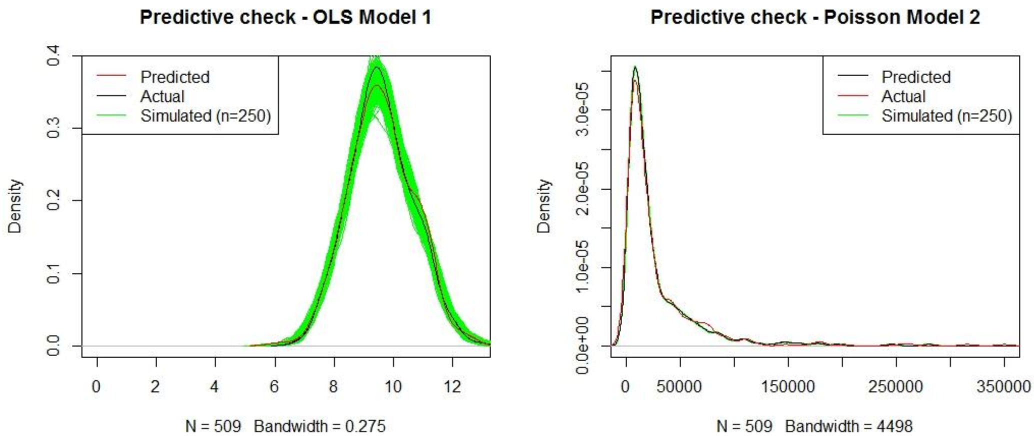


Figure 6. 8: Predictive check simulation of out-migration models

The predictive checking test in Figure 6.8 also produced a similar result to in-migration model. The simulated curve fits best with the Poisson model, but since the outcome of the Poisson model is over-dispersed, a negative binomial model is preferred.

According to Table 6.11, there are 12 explanatory variables that are significant in the negative binomial model: variables that increase the amount of out-migration are total population, percentage of children/adolescents, mature adult, those who attained tertiary education, tertiary industry workers; In contrast, variables that reduce the size of out-migration flows are all years, percentage of ethnic majority, unemployed people, and secondary industry worker.

More out-migration flows are estimated in areas that have more children/adolescents and mature adults probably because of parent-children related that out-migrate due to work commitment of the parent (Jali, 2009; UNESCO et al., 2012). Further, people who are more educated are normally inclined to out-migrate to search for jobs equal to their qualifications (Gallup, 1996). For example, jobs that require tertiary education (e.g., professionals, managers) are commonly located in larger settlements. More out-migration also happens in areas that offer more tertiary industry jobs. One of the main reasons is that these jobs can commonly be found anywhere in Malaysia (e.g., wholesale and retail trade, hotels and restaurants, transportation) so those who out-migrate are probably doing so due to changing employment or getting a new job in another area (Jali, 2009).

Surprisingly, less out-migration happens in areas that have a high unemployment rate. Economic theories and empirical evidence suggest that unemployment and migration have an inverse relationship (Rees et al., 1996). There are three possible reasons for this: 1) the lack of resources to move such as money or transportation; 2) the 2 percent sample size used may not be sufficient to represent total out-migration in Malaysia; and 3) other variables are controlled for in the model, hence reversing the unemployment effects. Furthermore, less out-migration also happens in areas that offer more secondary industry jobs because these jobs normally require workers to stay within the working area rather than migrating out to other regions. For example, manufacturing and construction activities are usually located in specific areas (e.g., cities) or areas near natural resources, requiring workers to stay in the vicinity (Jali, 2009).

Overall, even though all models (net migration, in-migration, and out-migration) are built using different regression approaches, the results display a similar pattern and agree with each other. The Poisson model captures more explanatory variables than the ordinary least squares model and negative binomial model. Further, there is a similar pattern of some negative or positive coefficients between the models. The negative binomial model is preferable since the ordinary least squares model and Poisson model must be rejected due to high uncertainty and over-dispersion issues.

Finally, the results of the aggregate flow models capture only a few determinants (e.g., the net migration model captures only one determinant). This is because this approach models only the aggregate flow to each place, rather than the full set of individual origin-destination flows. In addition, according to Rogers (1990), there is no such thing as net migrants because internal migration consists of people who in-migrate and out-migrate. Indeed, the next chapter on population and migration projections models out-migrant and in-migrant flows separately, for precisely the reasons explained by Rogers. Hence, the rest of this chapter focusses on modelling flows from origin to destination rather than aggregate and net migration flows.

6.6 Spatial interaction models

To build a full spatial interaction model of origin-destination flows, the following sections discuss the development of the model, starting with the unconstrained null model in Section 6.6.1 and followed by the doubly constrained model in Section 6.6.2, the origin-constrained model in Section 6.6.3, the destination-constrained model in Section 6.6.4, and the unconstrained model in Section 6.6.5. Finally, Section 6.6.6 maps out the spatial distribution of flow residuals of the unconstrained model to check for spatial autocorrelation. A detailed explanation of these models is provided explained in the respective sections.

6.6.1 Unconstrained null model

Table 6. 12: Relationship of flow to distance (actual values, exponential, and log)

Variables	Model 1		Model 2		Model 3	
	B	Sig	B	Sig	B	Sig
Intercept	273.496	0.000	271.067	0.000	1,188.989	0.000
d_{ij}	-0.192	0.000				
(d_{ij} /1000) exp			-56.663	0.000		
d_{ij} (log)					-171.017	0.000
R-squared		0.031		0.021		0.076

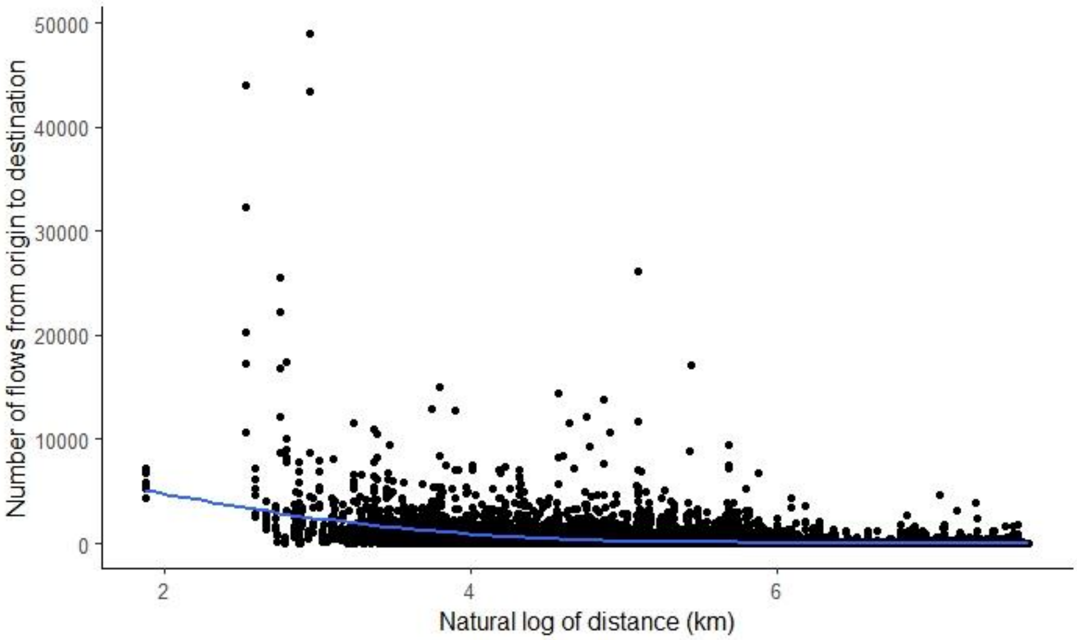


Figure 6. 9: Scatter plot of size of flow against distance of flow

The results of the unconstrained null model in Table 6.12 and the scatter plot in Figure 6.9 show a distance decay effect. Distance decay is defined as the effect of distance over spatial interactions, in this case the effect of distance over migration flows. The shorter the distance, the greater the estimated flow. This result agrees with Ravenstein's first migration law; most migrants move to a short distance rather than a long distance. Logging the distance attribute gives a better prediction (0.076) than the actual (0.031) and exponential values (0.021). However, the model fit (the R-squared value) is very low because flows cannot be explained by distance only. Furthermore, this model only observes distance decay for all years studied, 1970-2000. Therefore, the next step is to distinguish the distribution values by each year to identify whether distance decay varies over time.

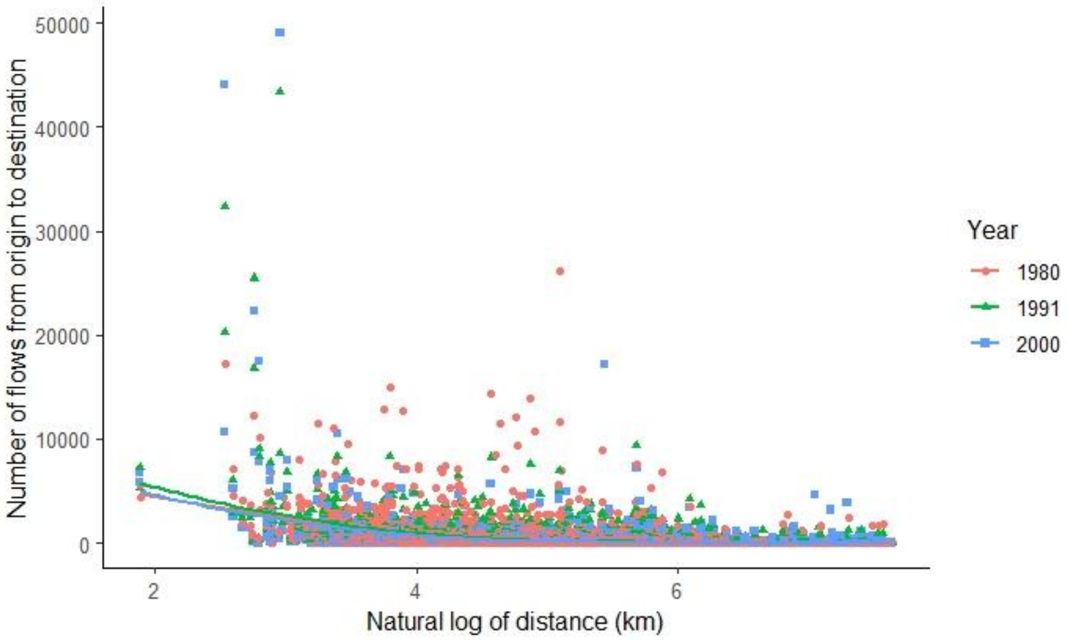


Figure 6. 10: Scatter plot of size of flow against distance of flow for each year

Based on Figure 6.10, the distribution values are similar for each year, meaning distance decay does not vary over time. The results from this analysis reflect the raw flows, taking no account of any origin or destination attributes, and portray a distance decay pattern as postulated in migration theory. In reality, flows from one place to another are mostly influenced by origin and destination attributes instead of distance. For example, people are willing to travel further to reach big cities such as Kuala Lumpur than small towns or rural villages. This shortcoming sets up the next set of models, which model the flows between origin-destination pairs.

6.6.2 Doubly constrained model

This section tests and selects the regression model that best explains flows by constraining the origin and destination. Table 6.13 shows the relationship of flows to the place of origin, destination, and distance. Similar to the aggregate migration modelling method, the results in Table 6.13 show six variant models built using different regression approaches. In general, the results for all models agree with each other. The distance variable has a negative coefficient, and logging the distance produces a better model fit (Model 3, Model 6, and Model 9 based on the R-squared value for ordinary least squares regression or AIC for Poisson and negative binomial) than models that use actual values and exponential values.

Table 6. 13: Relationship of flows to place of origin and destination, and distance

Variables	Ordinary least square						Poisson						Negative binomial					
	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6		Model 7		Model 8		Model 9	
	B	Sig	B	Sig	B	Sig	B	Sig	B	Sig	B	Sig	B	Sig	B	Sig	B	Sig
Intercept	248.00	0.00	263.91	0.00	1,121.45	0.00	5.91	0.00	7.18	0.00	10.72	0.00	5.64	0.00	5.75	0.00	7.64	0.00
origin**		-																
destination**																		
d_{ij}	-0.18	0.00					-0.00	0.00					-0.00	0.00				
exp (d_{ij} /1000)			-55.70	0.00					-1.70	0.00					-0.21	0.00		
d_{ij} (log)					-168.33	0.00					-1.20	0.00					-0.43	0.00
R-squared		0.127		0.121		0.166		-		-		-		-		-		-
AIC		767,523		767,849		765,302		7,770,313		9,441,261		5,364,720		568,993		573,294		551,293
Residual deviance		-		-		-		7,609,405		9,280,352		5,203,811		50,717		50,969		49,828
Degrees of freedom		-		-		-		48,532		48,532		48,532		48,532		48,532		48,532

Note:
** +100 to all $Flow_{ij}$ for negative binomial models (Model 7, Model 8, and Model 9) to obtain model convergence.*
***The coefficients of origin and destination dummy variables are excluded due to the long lists.*

To select the best regression model, predictive checking tests were conducted between the best models based on R-squared or AIC: Model 3 (ordinary least squares) and Model 6 (Poisson).

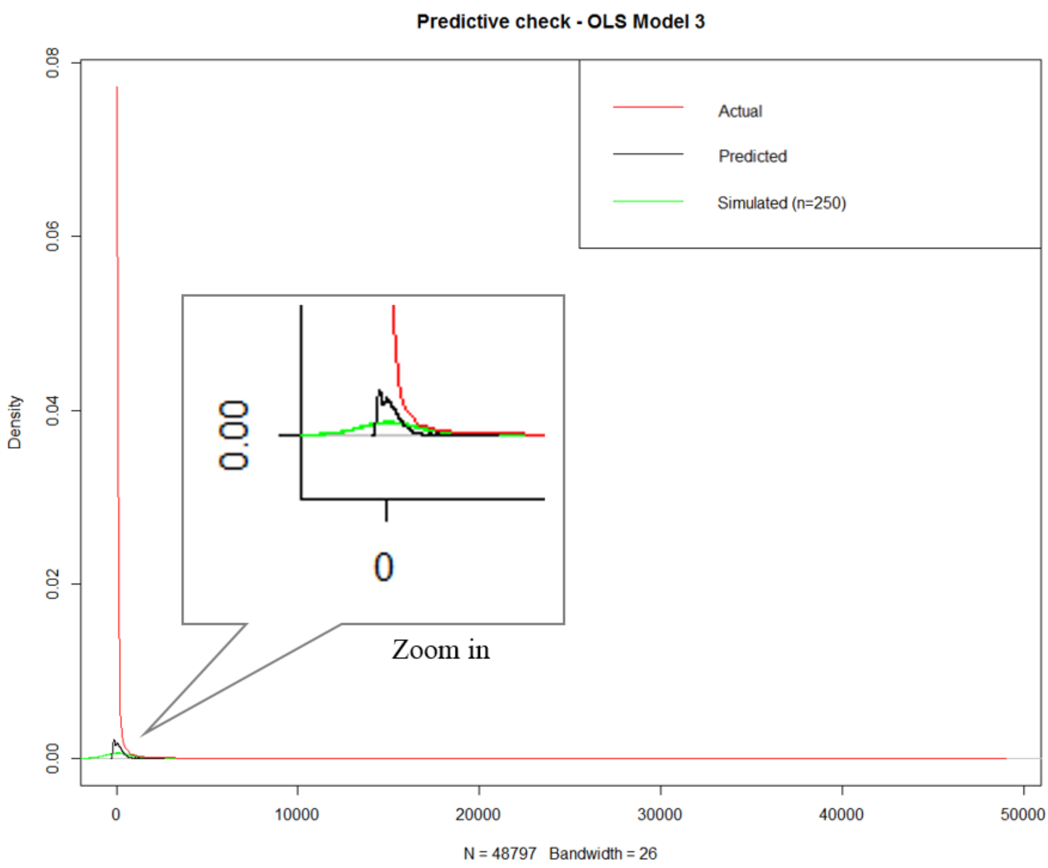


Figure 6. 11: Predictive check of Model 3

Figure 6.11 shows the results of the predictive checking test for Model 3. First, there is a major difference between the actual and predictive values or curve. Because ordinary least squares regression assumes the dependant variable follows a normal (Gaussian) distribution or bell shape, the distribution of predicted values portrays a similar pattern and contra to the actual values. Secondly, the simulated values or curve generated from random predictions mismatch the predicted values. This result shows ordinary least squares regression is not the best way to examine flows. For one, flows cannot be negative, and the actual distribution looks more like a Poisson distribution than a normal distribution. Other reasons are poor model estimation results in diverse predictions and large errors caused by a large amount of unexplained variation in the model.

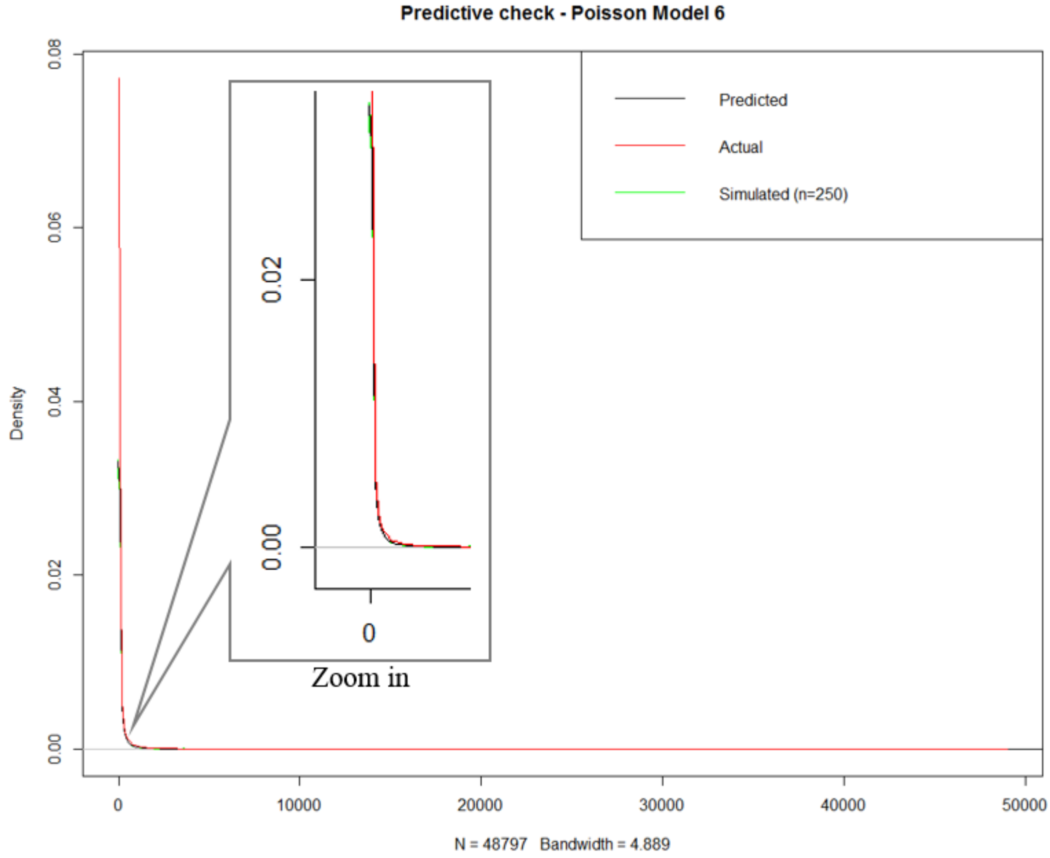


Figure 6. 12: Predictive check of Model 6

Figure 6.12 shows the results of the predictive checking test for Model 6, for which Poisson regression was adapted. As seen in the figure, the predicted and simulated values fit very closely to the actual values. This means the model offers a better prediction than Model 3. However, the outcome of Model 6 is over-dispersed; the difference between residual deviance and degrees of freedom is very large (520,381 compared to 48,532). Therefore, negative binomial regression is proposed to measure flows as well as over-dispersion. Model 9 appears to be the best model due to its lowest AIC, indicating it has a better model fit than Model 7 and Model 8. The over-dispersion is handled appropriately whereas the residual deviance and degrees of freedom are approximately similar (49,828 to 48,532).

From this point onwards, Model 9 or negative binomial regression is used to develop a more sophisticated model.

Table 6. 14: Relationship of flows to distance and years

Variables	Null model		Model constrained by years		Model by each year					
	Model 1		Model 2		Model 3 (1970-80)		Model 4 (1980-91)		Model 5 (1991-00)	
	B	Sig	B	Sig	B	Sig	B	Sig	B	Sig
Intercept	7.671	0.00	7.643	0.00	7.851	0.00	7.690	0.00	6.977	0.00
origin*										
destination*										
d_{ij} (log)	-0.429	0.00	-0.427	0.00	-0.423	0.00	-0.461	0.00	-1.075	0.00
year8091			-0.026	0.00						
year9100			-0.102	0.00						
AIC	551,293		550,928		170,570		183,483		187,919	

*Note: *The coefficients of origin and destination dummy variables are excluded due to the long lists.*

Table 6.14 shows five model variants built to help select which model best explains each year of the flow. Judging from the AIC, the flow is best explained by each year (Model 3, Model 4 and Model 5) rather than aggregating all years into one model (Model 2) or not counting years (Model 1). The main reason for this is that the fixed effects (origin and destination dummy variables) are different for every year. There are 119 origins and 127 destinations during 1970-1980, 127 origins and 131 destinations during 1980-1991, and 131 origins and 133 destinations during 1991-2000.

The next step is to incorporate more flow attributes and model the flows by year. The results for each model are depicted in Tables 6.15 to 6.17. Seven models are tested: Model 1 is tested to check if there is a ‘crossing the sea’ effect; Model 2 is tested to check if there is ‘change of settlement type’ effect; Model 3 is tested to check if the ‘crossing the sea’ effect persists once the distance of the flow has been taken into account; Model 4 is tested to check if the ‘change of settlement type’ effect persists once the distance of the flow has been taken into account; Model 5 is tested to check whether having taken account of flows distance, and the ‘crossing the sea’ effect, a ‘change of settlement type’ effect persists; Model 6 and Model 7 are tested to check for possible interactions between the various flow attributes.

Table 6. 15: Size of flow against attributes of flow for 1970-1980

Variables/ Models	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6		Model 7	
	B	Sig	B	Sig	B	Sig	B	Sig	B	Sig	B	Sig	B	Sig
Intercept	5.825	0.000	5.696	0.000	8.976	0.000	7.803	0.000	8.908	0.000	5.825	0.000	7.809	0.000
origin*														
destination*														
d_{ij} (log)					-0.666	0.000	-0.413	0.000	-0.652	0.000			-0.415	0.000
crossing_sea	-1.162	0.000			1.272	0.000			1.235	0.000	-1.162	0.000		
not_crossing_sea	**	**			**	**			**	**	**	**		
sub_sub			2.087	0.000			1.059	0.000	0.733	0.000			4.486	0.000
sub_core			2.019	0.000			0.955	0.000	0.490	0.010			-0.565	0.664
sub_other			**	**			**	**	**	**			**	**
other_core			**	**			**	**	**	**			**	**
other_sub			**	**			**	**	**	**			**	**
other_other			**	**			**	**	**	**			**	**
d_{ij} (log)*acrossing_sea											**	**		
d_{ij} (log)*not_crossing_sea											**	**		
d_{ij} (log)*sub_sub													-0.649	0.000
d_{ij} (log)*sub_core													-0.468	0.216
d_{ij} (log)*sub_other													0.195	0.000
d_{ij} (log)*other_core													-0.739	0.000
d_{ij} (log)*other_sub													-0.020	0.202
d_{ij} (log)*other_other													**	**
AIC		177,366		179,452		168,483		170,322		168,368		177,366		169,912

Note:
**The coefficients of origin and destination dummy variables are excluded due to the long lists.*
***The variables are removed due to multicollinearity.*

Table 6. 16: Size of flow against attributes of flow for 1980-1991

Variables/ Models	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6		Model 7	
	B	Sig	B	Sig	B	Sig	B	Sig	B	Sig	B	Sig	B	Sig
Intercept	5.377	0.000	5.217	0.000	8.620	0.000	7.640	0.000	8.551	0.000	5.377	0.000	7.645	0.000
origin*														
destination*														
d_{ij} (log)					-0.653	0.000	-0.451	0.000	-0.639	0.000			-0.451	0.000
crossing_sea	-1.439	0.000			1.065	0.000			1.024	0.000	-1.439	0.000		
not_crossing_sea	**	**			**	**			**	**	**	**		
core_sub			2,285	0.000			1.039	0.000	0.628	0.000			3.866	0.000
core_other			**	**			**	**	**	**			**	**
sub_sub			1.832	0.000			0.748	0.000	0.463	0.000			5.490	0.000
sub_core			2.456	0.000			1.226	0.000	0.801	0.010			5.533	0.000
sub_other			**	**			**	**	**	**			**	**
other_core			**	**			**	**	**	**			**	**
other_sub			**	**			**	**	**	**			**	**
other_other			**	**			**	**	**	**			**	**
d_{ij} (log)*crossing_sea											**	**		
d_{ij}(log)*not_crossing_sea											**	**		
d_{ij} (log)*core_sub													-1.385	0.000
d_{ij} (log)*core_other													-0.384	0.000
d_{ij} (log)*sub_sub													-1.113	0.000
d_{ij} (log)*sub_core													-1.892	0.000
d_{ij} (log)*sub_other													-0.003	0.793
d_{ij} (log)*other_core													-0.274	0.000
d_{ij} (log)*other_sub													0.116	0.000
d_{ij} (log)*other_other													**	**
AIC		194,168		198,096		181,339		183,169		181,218		194,168		182,821

*Note:*The coefficients of origin and destination dummy variables are excluded due to the long lists.*
***The variables are removed due to multicollinearity.*

Table 6. 17: Size of flow against attributes of flow for 1991-2000

Variables/ Models	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6		Model 7	
	B	Sig	B	Sig	B	Sig	B	Sig	B	Sig	B	Sig	B	Sig
Intercept	5.266	0.000	5.181	0.000	7.781	0.000	6.898	0.000	8.551	0.000	5.266	0.000	6.882	0.000
origin*														
destination*														
d_{ij} (log)					-0.501	0.000	-0.320	0.000	-0.473	0.000			-0.318	0.000
crossing_sea	-0.929	0.000			0.912	0.000			0.827	0.000	-0.929	0.000		
not_crossing_sea	**	**			**	**			**	**	**	**		
core_sub			2,771	0.000			1.930	0.000	1.599	0.000			3.999	0.000
core_other			**	**			**	**	**	**			**	**
sub_sub			1.947	0.000			1.192	0.000	0.955	0.000			6.349	0.000
sub_core			1.827	0.000			0.937	0.000	0.607	0.010			4.800	0.000
sub_other			**	**			**	**	**	**			**	**
other_core			**	**			**	**	**	**			**	**
other_sub			**	**			**	**	**	**			**	**
other_other			**	**			**	**	**	**			**	**
d_{ij} (log)*crossing_sea											**	**		
d_{ij}(log)*not_crossing_sea											**	**		
d_{ij} (log)*core_sub													-1.356	0.000
d_{ij} (log)*core_other													-0.366	0.000
d_{ij} (log)*sub_sub													-1.362	0.000
d_{ij} (log)*sub_core													-1.565	0.000
d_{ij} (log)*sub_other													0.054	0.000
d_{ij} (log)*other_core													-0.236	0.000
d_{ij} (log)*other_sub													-0.018	0.146
d_{ij} (log)*other_other													**	**
AIC		195,437		196,912		186,083		187,034		185,508		195,437		186,692

*Note:*The coefficients of origin and destination dummy variables are excluded due to the long lists.*
***The variables are removed due to multicollinearity.*

Tables 6.15 to 6.17 show consistent results across the years, with the coefficients displaying similar positive or negative coefficients. Model 5 has the best model fit for all years of flow compared to all other models because it has the lowest AIC. A point to note is that some flow attributes are removed in Model 5 due to multicollinearity (high correlation between explanatory variables). Multicollinearity exists due to origin and destination fixed effects. For example, the flow from core city (origin) to another settlement type (destination) is similar to the flow from another settlement type (origin) to core city (destination). Greater distance leads to smaller flows whereas crossing the sea and moving between settlement types are associated with larger flows.

The following section uses Model 5, separated for each year, to examine the influence of origin and destination socio-economic attributes on flows.

6.6.3 Origin-constrained model

An origin-constrained model correctly predicts the total number of flows from a given origin. However, it does not constrain the number of flows to each predicted destination. Instead, destination attributes that may explain the size of the flow to each destination are added to the model. As mentioned in Section 6.3, all destination attributes applied in the model are considered to be important based on theoretical reviews. For example, neoclassical migration theories such as Harris-Todaro's two-sector model assume migrants' decisions are made based on rational economic reasons such as wage and job opportunities. Further, the dualistic economy model highlights the absorption of surplus labour from the agricultural sector into the industrial sector in urban areas (Young, 2004).

Table 6.18 shows the result of origin-constrained model by each year. Technically, a positive coefficient means that if the destination attributes increase by 1 (e.g., percentage), then more flows are estimated. In contrast, a negative coefficient means that if the destination attributes increase, then fewer flows are estimated. In other words, more migrants are pulled by the destination attribute if the coefficient values are positive, while fewer migrants are pulled by the destination attributes if the values of the coefficient are negative.

Table 6. 18: Origin-constrained models

Variables	Model 1 (1970-80)		Model 2 (1980-91)		Model 3 (1991-00)	
	B	Sig	B	Sig	B	Sig
Intercept	10.280	0.000	7.584	0.000	6.199	0.000
origin*						
Flow attributes						
d_{ij}	-0.545	0.000	-0.583	0.000	-0.425	0.000
core_sub	**	**	0.503	0.001	1.606	0.000
core_other	**	**	***	***	***	***
sub_sub	0.952	0.000	0.328	0.000	0.883	0.000
sub_core	1.943	0.000	1.437	0.000	0.914	0.000
sub_other	***	***	***	***	***	***
other_core	1.237	0.000	0.519	0.000	0.207	0.000
other_sub	0.080	0.000	-0.209	0.000	-0.128	0.000
other_other	***	***	***	***	***	***
crossing_sea	0.502	0.000	0.729	0.000	0.555	0.000
not_crossing_sea	***	***	***	***	***	***
Destination attributes						
%_Total_population$_j$ (log)	0.224	0.000	0.255	0.000	0.217	0.000
%_Male population$_j$			0.004	0.012	0.016	0.000
%_Young_adult$_j$ (log for model 1)	-0.226	0.000	-0.008	0.000		
%_Middleaged_adult$_j$	-0.024	0.000			0.008	0.004
%_Mature_adult$_j$	-0.051	0.000			-0.015	0.000
%_Elderly$_j$	0.017	0.000			0.007	0.044
%_Workingaged_adult$_j$			0.007	0.000		
%_Ethnic_majority$_j$					0.002	0.000
%_Married$_j$ (log for model 1)	0.332	0.001			0.009	0.000
%_Widowed_Separated$_j$					-0.012	0.048
%_No_formal_education$_j$			0.006	0.000		
%_Attained_primary_education$_j$	-0.005	0.000				
%_Attained_secondary_education$_j$			0.032	0.000	-0.017	0.004
%_Attained_tertiary_education$_j$ (sqrt for model 1)	-0.285	0.000	-0.018	0.004	0.123	0.000
%_Unemployed$_j$ (log for model 3)	-0.058	0.000	-0.004	0.060	-0.032	0.004
%_Inactive$_j$	-0.012	0.000			-0.005	0.000
%_Lowpaid_workers$_j$ (sq)	0.000	0.000				
%_Mediumpaid_workers$_j$			0.009	0.000		
%_Highpaid_workers$_j$	0.022	0.000			0.014	0.000
%_Secondary_industry_workers$_j$ (sqrt for model 1)	0.038	0.000	0.001	0.038		
%_Tertiary_industry_workers$_j$	-0.004	0.002				

AIC	171,330	182,143	187,740

Note:
** The coefficients of origin dummy variables are excluded due to the long lists.*
*** There is no core_sub or core_other flow because the core city of capital metropolitan areas did not exist in 1970.*
**** These variables are removed due to multicollinearity.*

1970 to 1980 – Destinations containing a higher percentage of the elderly attract more flows, while destinations with a higher percentage of younger people attract smaller flows. This result is similar as the finding in Chapter 5 that elderly migrants have greater mobility compared to the younger population, likely due to retirement. More flows are also estimated in areas with a percentage of the population that is married. Normally, spouses, especially wives, will follow their spouse to live near their workplace. According to UNICEF (2012), following family and marital status are important factors of internal migration in Malaysia. On the contrary, destinations containing a higher percentage of those who obtained primary or tertiary education discourage flows because it is likely that people have secured a job in the area in which they live using their qualifications.

In terms of socio-economic factors, destinations with high unemployment and a large proportion of the population that is economically inactive population attract less flow. In classical migration theory, migration is mainly driven by economic motivation; normally, people migrate to areas that have high employment rather than to areas that have high unemployment (Harris & Todaro, 1970). Further, destinations that offer more low-paid jobs, high-paid jobs, and secondary industry jobs attract more flows. Low-paid jobs consist of skilled agricultural and fishery workers, crafts and related trades workers, plant and machine operators and assemblers, and other basic occupations. High-paid jobs consist of legislators, senior officials, managers, and professionals. Secondary industry job consist of manufacturing and construction activities. Low-paid jobs are commonly found in rural areas while high-paid jobs and secondary industry jobs are mostly located in cities (refer to Chapter 4, Section 4.5.5). The disparity in the location of these jobs (urban vs. rural) indicates that flows are influenced not only by destinations that offer higher wages or more job opportunities (e.g., cities) but also by rural-based jobs. The influence of rural-based job is probably due to rural settlement schemes (e.g., FELDA, DARA, KEJORA) imposed by the Malaysian government during the 1970s to improve the economics of rural communities and reduce mass rural-urban migration (Abdullah, 2012). Furthermore, destinations that have a higher percentage of tertiary industry jobs attract less flow. Tertiary industry jobs are a combination of all services, such as electricity, gas and water, wholesale and retail trade, hotels and restaurants, and transportation. This type of job is commonly found in all locations, so it is unsurprising that they have little influence on flows.

1980 to 1991 – Unlike the previous period, flows in this period were influenced by different push factors. More flows are estimated in destinations with a higher percentage of males. This result may relate to the Ravenstein's migration law that states men are more likely to migrate, especially for long distances, than women. The imbalanced growth of urban development between regions in Malaysia often requires travel for long distances or even across the sea. This is because West Malaysia is more developed and urbanised than East Malaysia, and the two regions are separated by the sea. Furthermore, more flow is estimated in destinations that have a larger working-age population, which is typical in every country. In terms of academic achievement, destinations that have a larger percentage of those who obtained tertiary education attract less flow.

In terms of socio-economic factors, as expected, there is less flow for destinations that have high unemployment. In contrast, destinations that offer more medium-paid jobs and secondary industry jobs attract more flow. Middle-paid jobs consist of technicians, associates, professionals, clerks, service workers, and shop and market sellers. The increased attractiveness of destinations that have more of these jobs reflect the economic transition from primary to multi-sector sector industries in the early 1980s. The manufacturing sector and modern services generally grow substantially and become centralized in the vicinity of cities (Abdullah, 2003).

1991 to 2000 – Similar to the previous periods, destinations that have a larger percentage of males and married individuals attract more flow. The only difference in this period is that there is more flow in areas that have more middle-aged adults, ethnic majority members, and those who achieved tertiary education. In contrast, less flow is estimated in areas that have more widowed individuals. Based on the findings in Chapter 5, the migration of middle-aged adults during this period was mainly due to urban-urban migration that was the result of the decentralisation of urban development, urban sprawl, and rapid suburbanisation. The migration of ethnic majority members was mostly due to the attractiveness of cities due to an improved economic situation. Historically, the ethnic majority (also known as *Bumiputera*) was redistributed by the British colonizer in rural areas to focus on farming and agricultural activities while minorities focused on trade and business activities. As the country achieved independence, the Malaysian government imposed various policies and strategies beginning in 1970 (e.g., National Economic Policy) to restructure communities and eradicate poverty, especially among the ethnic majority. In terms of academic achievement, as more universities were established and produced more graduates (i.e., tertiary education), graduates often had to migrate to other areas to find jobs that matched their qualifications. This relates to the positive relationship between flows and high-paid; areas that have a higher percentage of high-paid jobs attract more flow. As mentioned previously, high-paid jobs consist of professional and managerial positions that require a tertiary education qualification and are normally found mostly in cities.

6.6.4 Destination-constrained model

In contrast to the origin-constrained model, the destination-constrained model correctly predicts the total amount of flow to a given destination and makes no constraint on the flow from each predicted origin. Instead, origin attributes that could explain the size of the flow to each origin are considered. Table 6.19 shows the results of destination-constrained model by each year.

Table 6. 19: Destination-constrained models

Variables	Model 1 (1970-80)		Model 2 (1980-91)		Model 3 (1991-00)	
	B	Sig	B	Sig	B	Sig
Intercept	0.462	0.428	7.227	0.000	7.194	0.000
destination*						
Flow attributes						
d_{ij}	-0.615	0.000	-0.557	0.000	-0.427	0.000
core_sub	**	**	1.315	0.000	2.173	0.000
core_other	**	**	0.510	0.000	0.482	0.000
sub_sub	0.389	0.000	0.652	0.000	0.865	0.000
sub_core	0.184	0.358	1.129	0.000	0.514	0.000
sub_other	-0.356	0.000	***	***	-0.157	0.000
other_core	***	***	***	***	***	***
other_sub	***	***	***	***	***	***
other_other	***	***	***	***	***	***
crossing_sea	0.994	0.000	0.523	0.000	0.573	0.000
not_crossing_sea	***	***	***	***	***	***
Origin attributes						
%_Total_population$_i$ (log)	0.336	0.000	0.211	0.000	0.251	0.000
%_Male_population$_i$			-0.007	0.000		
%_Young_adult$_i$ (log)	0.819	0.000				
%_Middleaged_adult$_i$			-0.018	0.000	0.011	0.000
%_Mature_adult$_i$	0.023	0.000	-0.034	0.000	0.017	0.000
%_Elderly$_i$	0.061	0.000				
%_Workingaged_adult$_i$	-0.019	0.000				
%_Ethnic_majority$_i$			0.002	0.000	0.002	0.000
%_Single$_i$	0.072	0.000				
%_Married$_i$ (log for model 2)	0.060	0.000	0.955	0.000		
%_Widowed_Seperated$_i$					-0.027	0.000
%_Attained_primary_education$_i$	-0.020	0.000				
%_Attained_secondary_education$_i$ (sqrt)	-0.117	0.000	0.028	0.026		
%_Attained_tertiary_education$_i$ (sqrt for model 1 and 2)	-0.143	0.000	-0.132	0.000	-0.017	0.003
%_Employed$_i$	0.002	0.045				
%_Unemployed$_i$	-0.026	0.000	-0.040	0.000		
%_Inactive$_i$			-0.005	0.000	-0.003	0.000
%_Lowpaid_workers$_i$ (sq)			0.000	0.000		
%_Mediumpaid_workers$_i$					0.005	0.000
%_Highpaid_workers$_i$	0.010	0.000	0.012	0.000	0.010	0.000
%_Secondary_industry_workers$_i$ (sqrt)			0.041	0.000		
%_Tertiary_industry_workers$_i$ (log for model 1)	0.055	0.002	-0.012	0.000	-0.003	0.001
AIC		169,901		183,796		186,757

Note:
** The coefficients of destination dummy variables are excluded due to the long lists.*

*** There is no core_sub or core_other flow because the core city of capital metropolitan areas did not exist in 1970.*
**** The variables are removed due to multicollinearity.*

Similar to the origin-constrained model, each destination-constrained model captured different explanatory variables. Despite the difference, the coefficient of some variables produced similar coefficients (positive or negative), meaning the results are consistent for certain years. Contrary to origin-constrained models, the results for destination-constrained models can be interpreted differently; more migrants are pushed out by attribute at origin if the coefficient values are positive, while fewer migrants are pushed out by attributes at origin if the coefficient values are negative.

1970 to 1980 – Origins that have a higher percentage of young adults, mature adults, and the elderly encouraged more flow. In other words, more out-migration flow is estimated from origins that have a higher percentage of these age groups. Normally, these groups comprise people who are actively searching for jobs in other areas, who seek to migrate because their workplace is located in another place, or who seek to migrate due to retirement. In contrast, origins that have a higher percentage of the working-age population discouraged flows. This is likely due to the aggregation model effect. The working-age population (ages 15 to 64) comprises all adult groups (young adults, middle-aged adults, and mature adults) and some elderly. Because origins that have a higher percentage of young adults, mature adults, and the elderly encourage more flow, removing origins with higher percentages of middle-aged adults from the model resulted in contradictory results for origins with higher percentages of the working-age population.

Furthermore, high percentages of the population that are unmarried or married populations are push factors. Those who are unmarried are commonly people who have just finished school or graduated from university and are migrating to find jobs. As explained in the previous section, marriage or following family/a spouse is an important factor of migration in Malaysia. In terms of educational achievement, origins that have a higher percentage of people who obtained primary, secondary, and tertiary levels of education discourage flows. This result contradicts the theoretical perspective, whereby the more educated the population, the more migration is expected. A possible reason for this is that these people are unable to find or secure a job (this could relate to less flow estimated from origins that have high unemployment) or have already secured a job in the place where they live. Finally, origins that have a higher percentage of tertiary industry jobs encourage more flow. Tertiary industry jobs are those that provide services (e.g., restaurants, retail, and trade) and are available in nearly all locations. People may be pushed away from origins that offer this type of job because they are searching for similar jobs or jobs in another industry in another area.

1980 to 1991 – Attributes that induced more flows are origins that have a higher percentage of the ethnic majority, married population, and people that attained secondary education. Historically, most ethnic majority, or *Bumiputera*, settled in smaller towns and rural areas. However, due to racial tensions in 1969, the Malaysian government implemented the National Economic Policy to reduce ethnic segregation by encouraging the *Bumiputera* to become involved in trading and commercial activities in cities. In contrast, attributes that discourage flows are origins that have a higher percentage of males, middle-aged adults, mature adults, and people that have attained tertiary education. The reason for this is that people in these groups are likely bound to a job or have family commitments in the places where they live and are hence less likely to migrate. This relates to less flow from origins that have a higher percentage of the population that is economically inactive (e.g., housewives, children below working age, students). Further, less flow is estimated from origins that have a higher percentage of the population that has attained tertiary education because these people may be unemployed or not actively looking for a job. This relates to fewer flows from origins that have high unemployment.

In terms of socio-economic factors, attributes that caused more flow are origins that have a higher percentage of low-paid jobs, high-paid jobs, and secondary industry jobs. In contrast, origins with a higher percentage of tertiary industry jobs cause less flow. Generally, more people tend to migrate from an origin to search for similar or better paying jobs in other areas. This result relates to the origin-constrained model whereby destinations with secondary industry jobs and medium-paid jobs attract more flow.

1991 to 2000 – While some variables display similar explanatory variables and coefficient (origins with a high percentage of the ethnic majority, people who attained tertiary education, high-paid jobs, and tertiary industry jobs) as in the previous periods, the model also captures other variables (origins with a high percentage of middle-aged adults, mature adults, widowed, economically inactive individuals, and medium-paid jobs). More flow was estimated from origins that have higher percentages of middle-aged adults and mature adults. This result is highly related to more flow from origins that have a higher percentage of economically inactive individuals. Adults who are economically inactive are likely family-oriented and may migrate from an origin when following their family. Additionally, less flow was estimated in origins that have more widowed individuals. Unlike the married population, many of those who are widowed have no reason to migrate out of the place in which they live. Finally, more medium-paid jobs in an origin encourages more flow. Similar to the previous reasons, those who out-migrate most probably do so due to limited job vacancies (generally for government servants such as technicians, associates, professionals) or are looking for better jobs. Accordingly, destinations with a higher percentage of high-paid jobs attract more flow based on the origin-constrained model.

6.6.5 Unconstrained model

Table 6. 20: Unconstrained models

Variables	Model 1 (1970-80)		Model 2 (1980-91)		Model 3 (1991-00)	
	B	Sig	B	Sig	B	Sig
Intercept	7.490	0.000	4.813	0.000	5.958	0.000
Flow attributes						
d_{ij}	-0.539	0.000	-0.549	0.000	-0.407	0.000
core_sub	*	*	1.064	0.000	2.124	0.000
core_other	*	*	0.509	0.000	0.475	0.000
sub_sub	0.564	0.000	0.449	0.000	0.739	0.000
sub_core	1.589	0.000	1.687	0.000	0.763	0.000
sub_other	-0.361	0.000	0.080	0.000	-0.169	0.000
other_core	1.229	0.000	0.508	0.000	0.217	0.000
other_sub	0.081	0.000	-0.233	0.000	-0.129	0.000
other_other	***	***	***	***	***	***
crossing_sea	0.424	0.000	0.495	0.000	0.464	0.000
not_crossing_sea	***	***	***	***	***	***
Origin attributes						
%_Total_population$_i$ (log)	0.285	0.000	0.230	0.000	0.252	0.000
%_Young_adult$_i$ (log)	0.354	0.000				
%_Middleaged_adult$_i$			-0.017	0.000	0.010	0.000
%_Mature_adult$_i$			-0.034	0.000	0.019	0.000
%_Elderly$_i$	0.040	0.000				
%_Ethnic_majority$_i$			0.002	0.000	0.002	0.000
%_Single$_i$	0.017	0.000				
%_Married$_i$ (log)			0.895	0.000		
%_Widowed_Seperated$_i$					-0.035	0.000
%_Attained_primary_education$_i$	-0.013	0.000				
%_Attained_secondary_education$_i$ (sqrt)	-0.089	0.000				
%_Attained_tertiary_education$_i$ (sqrt for model 1 and 2)	-0.062	0.000	-0.088	0.000	-0.016	0.011
%_Unemployed$_i$			-0.038	0.000		
%_Inactive$_i$			-0.005	0.000	-0.006	0.000
%_Mediumpaid_workers$_i$					0.002	0.002
%_Highpaid_workers$_i$	0.006	0.000	0.017	0.000	0.007	0.000
%_Secondary_industry_workers$_i$ (sqrt)			0.048	0.000		
%_Tertiary_industry_workers$_i$			0.002	0.000		
Destination attributes						
%_Total_population$_j$ (log)	0.231	0.000	0.259	0.000	0.221	0.000
%_Male_population$_j$					0.016	0.000
%_Young_adult$_j$ (log for model 1)	-0.193	0.001	-0.003	0.010		
%_Middleaged_adult$_j$	-0.020	0.000			0.008	0.005
%_Mature_adult$_j$	-0.050	0.000			-0.017	0.000

%_Elderly$_j$	0.014	0.000			0.007	0.037
%_Workingaged_adult$_j$			0.008	0.000		
%_Ethnic_majority$_j$					0.002	0.000
%_Married$_j$ (log for model 1)	0.297	0.006			0.012	0.000
%_Widowed_Seperated$_j$					-0.014	0.020
%_No_formal_education$_j$			0.008	0.000		
%_Attained_primary_education$_j$	-0.005	0.001				
%_Attained_secondary_education$_j$			0.012	0.012	-0.022	0.000
%_Attained_tertiary_education$_j$ (sqrt for model 1)	-0.244	0.000			0.120	0.000
%_Unemployed$_j$	-0.056	0.000				
%_Inactive$_j$	-0.011	0.000			-0.005	0.000
%_Mediumpaid_workers$_j$			0.009	0.000		
%_Highpaid_workers$_j$	0.022	0.000			0.014	0.000
%_Secondary_industry_workers$_j$ (sqrt)	0.040	0.000				
%_Tertiary_industry_workers$_j$	0.002	0.002				
AIC		172,493		184,476		188,891

Note:
** There is no core_sub and core_other flows because capital metropolitan cores were not established in 1970.*

Table 6.20 shows the final spatial interaction model, which is also known as an unconstrained model. As indicated by the name, this model places no constraints on the place of origin or destination to estimate flows. In fact, the flows are explained by the explanatory variables captured by the origin-constrained model and destination-constrained model. However, not all variables from previous models are included due to multicollinearity. The results are summarised here since they have already been explained in previous sections (Sections 6.6.3 and Section 6.6.4). Because the previous sections have already explained the attributes for each period, this section summarises the results by explaining only factors that attract more flow at destinations and factors that push more flow from origins.

Pull factors (destination attributes) – In terms of age groups, destinations with more adults, elderly, and working adults attract more flow. These results resemble those in Chapter 5 results, with older adults and elderly migrants having greater mobility than younger migrants. A common reason for this is retirement and changing workplaces. These results are surprising because young adults typically have the greatest mobility, which then declines with increasing age (Bernard et al., 2014). A possible reason for this is the decline in the number of persons of active migrant age (young adults) caused by a continuous decline in fertility since the

1960s. Further, more flows are estimated for destinations that has a higher percentage of the population that is married. As explained, following family and marital status are known to be primary reasons for internal migration in Malaysia (UNESCO et al., 2012). In the final period, destinations with more ethnic majority members attract more flow, possibly to cities. This relates to the results in Chapter 5, with the ethnic majority seeing larger growth than ethnic minorities in cities from 1980 until 2010, possibly due to the positive impact of national economic policies to encourage more flow among them. In terms of educational achievement, destinations with a more educated population (those who attained secondary and tertiary education) have attracted more flow since the second period. This results show the growing importance of education in the internal migration process.

For socio-economic factors, destinations that offer more high-paid and medium-paid jobs attract more flow. This is a classical migration pattern and is similar to Harris and Todaro's (1970) assumption that there is more migration into areas that offer higher wages or better jobs. Further, the increased attractiveness of destinations that offer more medium-paid and secondary industry jobs reflects the economic transition from primary to multi-sector sector industries in the early 1980s. The manufacturing sector and modern services grew substantially and became centralized in the vicinity of cities in this period (Abdullah, 2003).

Push factors (origin attributes) – In terms of age groups, origins with more adults (young, middle-aged, or mature adults) have more outflow. Furthermore, origins with more married people had more outflow in the second period, possibly because of following partners/spouses. However, in the first period, origins with a higher percentage of single individuals had more outflow. This possibly relates to more from origins with more young adults (as mentioned earlier) that are leaving school, enrolled in higher education, or entering the labour force. Further, origins with a higher percentage of the ethnic majority have more outflow. Similar to the reason stated before, this possibly relates to high growth of the ethnic majority population in all cities and encouragement from policies implemented by the Malaysian government.

In terms of socio-economic factors, for all periods, origins with more high-paid jobs consistently had more outflow. This is similar to the finding regarding origins with more middle-paid jobs/secondary industry jobs/tertiary industry job since the second period. A common reason this outflow is people seeking a similar or different type of job (see previous discussion about the jobs that attract more flow at a destination).

6.6.6 Spatial pattern of models' residual values

This section maps out models' residual values to examine spatial variation between flow sizes for each model (see Figures 6.13 to 6.15). Blue areas indicate the observed value (flow) is overpredicted while red areas indicates underpredicted. Further, the darker the colour, the more over- or underpredicted.

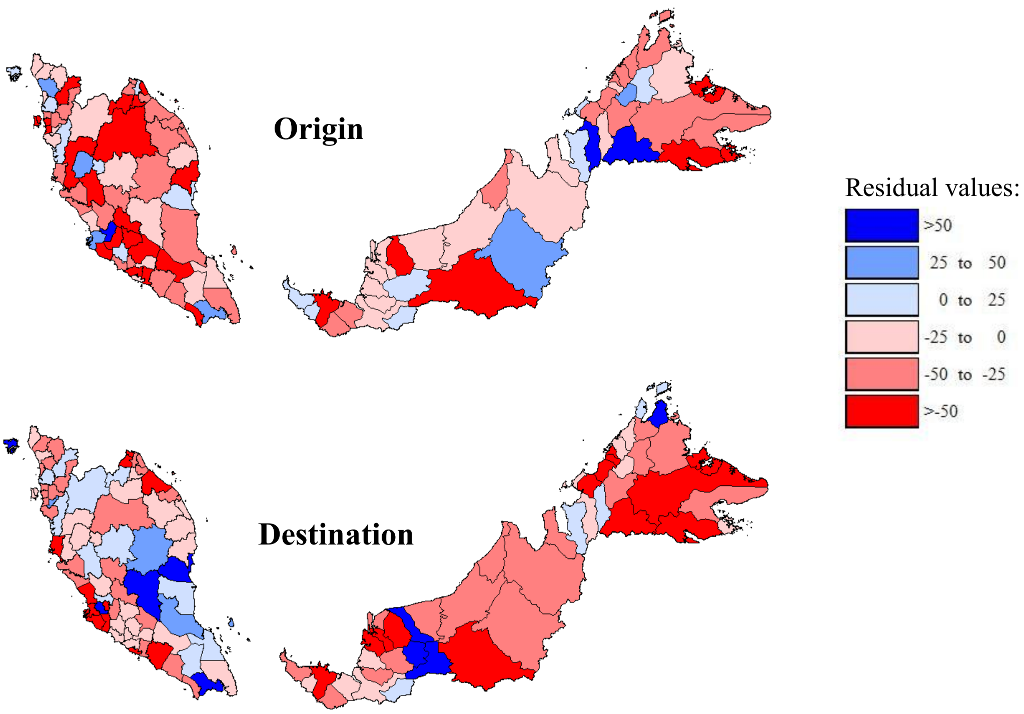


Figure 6. 13: Residual maps by place of origin and destination of flow, 1970-1980

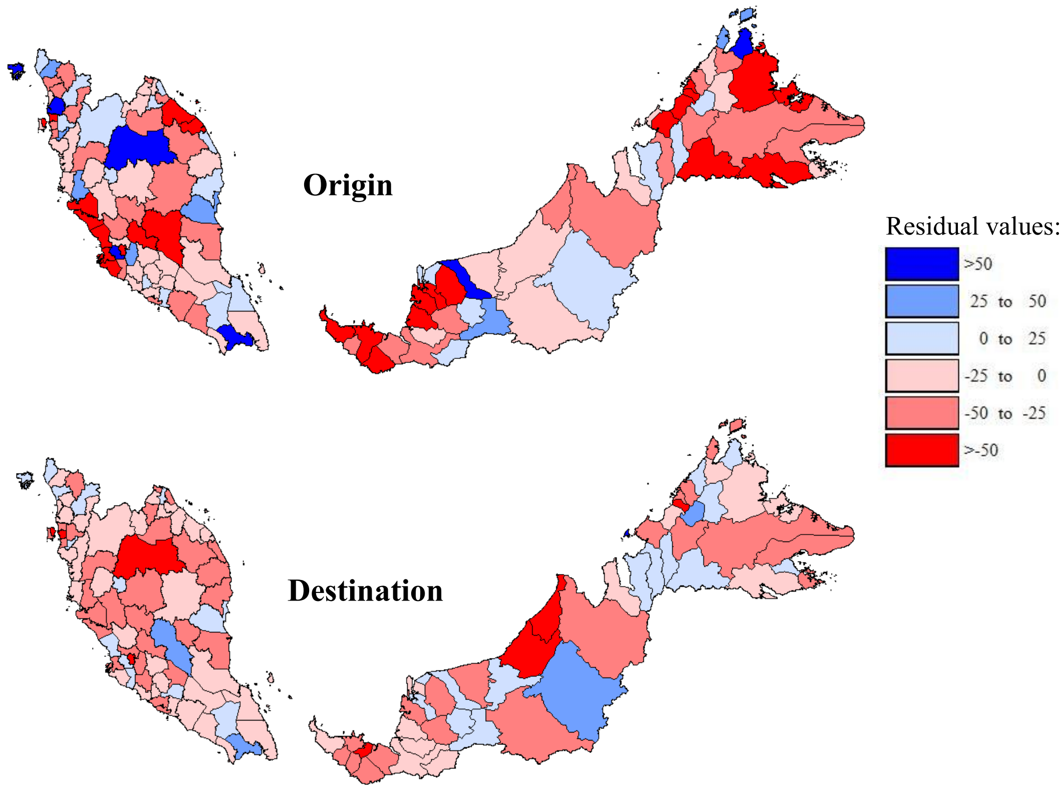


Figure 6. 14: Residual maps by place of origin and destination of flow, 1980-1991

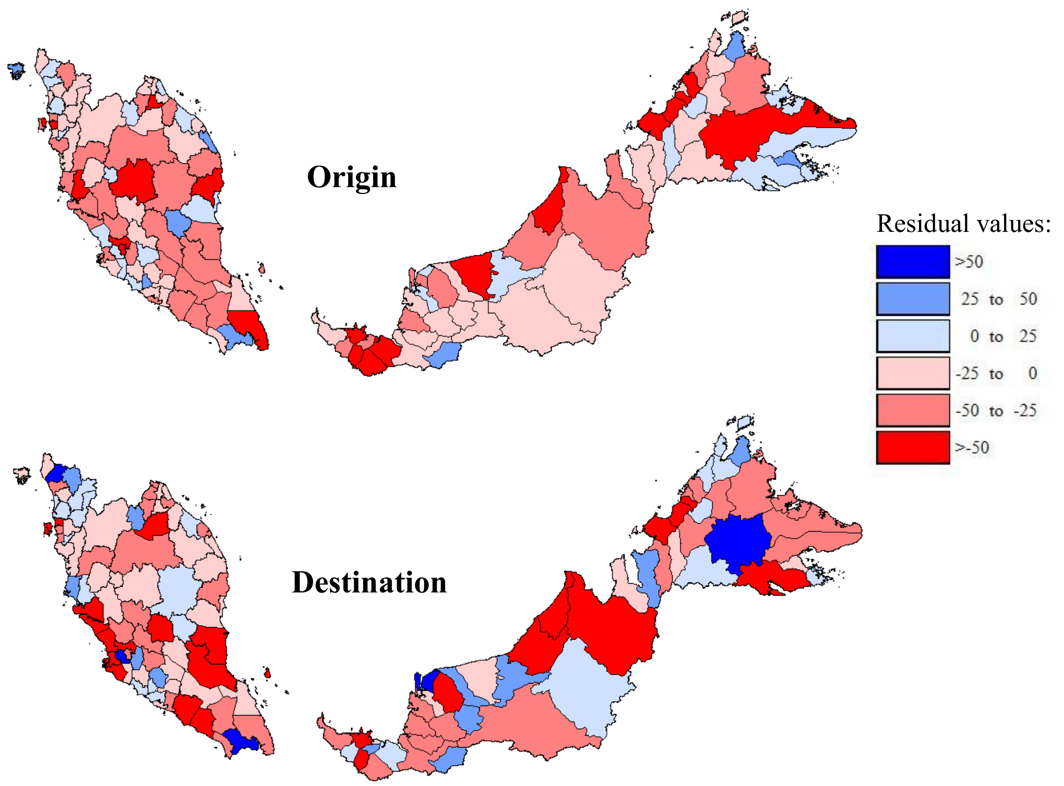


Figure 6. 15: Residual maps by place of origin and destination of flow, 1991-2000

Figures 6.14 to 6.16 show the residual maps by origin and destination for each period/ model. Visually, there is no specific or obvious spatial pattern for origin or destination maps for each period/model. To statistically examine the residual maps, spatial autocorrelation (Global Moran's I) is a good tool to measure the spatial correlation between neighbouring units (districts). The null hypothesis of spatial autocorrelation means the observed pattern (flow residual) is randomly distributed. In other words, the spatial processes promoting the observed pattern are by random chance. If the Moran's index is near 1 or -1 and is statistically significant (p-value less than 0.05), then the null hypothesis can be rejected. Table 6.11 shows the results of the spatial autocorrelation test of the models.

Table 6. 21: Spatial autocorrelation test

Model / Year	Residual by spatial unit	Spatial autocorrelation	
		Moran's Index	P-value
Flow 1970-80	Place of origin	0.010	0.657
	Place of destination	0.050	0.149
Flow 1980-91	Place of origin	0.022	0.455
	Place of destination	0.024	0.583
Flow 1991-00	Place of origin	0.054	0.116
	Place of destination	0.011	0.621

Table 6.19 shows all residual maps have a very small Moran's index (not close to 1 or -1) and the p-values are not statistically significant. This result confirms the visual assessment that the spatial distribution of the flow residual is the result of random spatial processes instead of clustering or an over-dispersed pattern. One of the reasons for this is the explanation of flows by multiple explanatory variables (e.g., age group, marital status, occupations), where the values are not similar between neighbouring districts. Another possible explanation is that the decision to migrate involves individual decision-making instead of an aggregate decision by migrants, hence there is no clustering pattern (Champion et al., 1998).

6.7 Conclusion

There are two main reasons spatial interaction models are preferable to examine the determinants of migration: 1) the models can identify the best set of factors influencing migration flows; and 2) the data structure is aggregated (e.g., net, in-, and out-migration of a district; and total flow from origin to destination by district), which fits the criteria of the model. To test the spatial interaction models comprehensively, two modelling approaches are proposed: 1) modelling total migration flow; and 2) modelling origin-destination flow. The total flow models capture only few determinants (e.g., the net migration model captures only one determinant) because this approach models only aggregate flows to each place, rather than the full set of individual origin-destination flows. Furthermore, according to Rogers (1990), there is no such thing as net migrations as migration consists of those who in-migrate and out-migrate. On the contrary, the origin-destination flow model can explain migration more clearly than the total migration flow models. The final model (unconstrained model) consists of the best sets of flow attributes (i.e., distance, crossing the sea, and movement between settlements), the best way to explain period of flow (by each period), and the best sets of determinants that propel the migrants from origins and attract them to destinations.

In conclusion, there is a clear distance effect, mediated by both crossing the sea between East and West Malaysia and flows between settlement types. The key drivers of flow (push factors) from origins are more adults (young, middle-aged, and mature adults), single individuals, married individuals, ethnic majority, high-paid jobs, middle-paid jobs, secondary industry jobs, and tertiary industry jobs. The origins with a higher percentage of the population that is married have more outflow, possibly due to following a partner/spouse. However, origins with a higher percentage of the population that is single also encourage more outflow. This relates to more flow from origins with more young adults leaving school, entering higher education, or entering the labour force. Further, origins with more ethnic majority have more outflow. This relates to high population growth of the ethnic majority in all cities and policies implemented by the Malaysian government. Finally, the increase of flows from origins that have higher percentages of multiple types of jobs is due to people seeking a similar or different type of job (see the discussion on jobs that pull more flow in the next paragraph).

The key drivers of flow (pull factors) from destinations are more adults, elderly, working adults, married individuals, ethnic majority, educated individuals (secondary and tertiary education), high-paid jobs, medium-paid jobs, secondary industry jobs, and tertiary industry jobs. These results are similar to those in Chapter 5, with older adults and elderly migrants having greater mobility than younger migrants. This is surprising because young adults typically have the greatest mobility, which then declines with increasing age (Bernard et al., 2014). A possible reason for this is the decline in the number of people of active migrant age (young adults) due to the continuous decline in fertility since the 1960s. Further, as previously explained, following family and marital status are known to be the primary reasons for internal migration in Malaysia (UNESCO et al., 2012). The increase of flows to areas that are attractive in terms of employment prospects agrees with classical migration theory that states there is more migration into areas that offer higher wages and better jobs (Harris & Todaro, 1970). Further, the increased attractiveness of destinations that offer more medium-paid and secondary industry jobs reflects the economic transition from primary to multi-sector sector industries in the early 1980s. The manufacturing sector and modern services grew substantially grew and became centralized in the vicinity of cities in this period (Abdullah, 2003).

In terms of the spatial distribution of flow residuals, the results show random spatial processes instead of clustering or an over-dispersed pattern. One of the reasons for this is that flows are explained by multiple explanatory variables (e.g., age group, marital status, occupations) with different values for neighbouring districts. Another possible explanation is that the decision to migrate involves individual decision-making instead of migrants' aggregate decisions, and hence there is no clustering pattern (Champion et al., 1998).

Chapter 7

Future population growth, internal migration, and urbanisation in Malaysia, 2015-2040

7.1 Introduction

Previous studies on differential urbanisation show Finland is the only country that has completed the first urbanisation cycle by undergoing counterurbanisation while other countries are still moving towards this. However, there are arguments that developing countries may not follow the urbanisation pattern assumed in differential urbanisation theory (Gedik, 2003; Mookherjee, 2003). This is due to the complex nature of developing countries in terms of historical, social, economic, and cultural conditions. This chapter tackles this question by predicting urbanisation trends in Malaysia trough population and migration projections from 2010 to 2040.

Countries, including Malaysia, produce their own projection statistics. The Malaysia Department of Statistics has projected the future population for the period of 2010-2040 at the national and state levels. This focus of this chapter, however, is on examining future population and migration change by settlement type instead of national or state levels, so that the projections can be related to differential urbanisation theory. Before implementing these new projections, it is important to review existing theories and models for population projection (Section 7.2). The next section, Section 7.3, introduces the data and methods used for the projections. The cohort-component model, combined with the iterative proportion fitting method, was chosen as the primary method. Furthermore, projections were only for Malaysian citizens (non-Malaysian citizens were excluded) due to several limitations. To check the quality of the projections, they are compared in Section 7.4 to the official projections by observing differences and similarities. This chapter's projections match the official national projections and are fairly similar to the official state projections. This is followed by Section 7.5, which examines future population change by settlement type. Capital metropolitan suburban areas is the only settlement type that shows significant change while other settlement types follow population patterns for the previous period (1980-2010). Population change results from either natural population growth (also known as natural increase or decrease) or migration. These aspects of population change are examined in Section 7.6 and Section 7.7, respectively. The results show that future population dynamics in Malaysia will be driven by natural increase (more births than deaths) and rapid urban-urban and rural-urban migration. To identify which population components (natural increase and migration) have the greatest influence, Section 7.8 and Section 7.9 examine individual components contributions to population change and the overall population, respectively. Perhaps surprisingly, natural population increase has a greater influence than migration. This is a rare situation because migration commonly has a greater impact on population change. Finally, Section 7.10 concludes all key findings in this chapter. Until at least 2040, based on differential urbanisation theory, Malaysia is predicted to remain in the final urbanisation stage, the APC stage, with population growth and migration flows to capital metropolitan suburban areas remaining dominant.

7.2 Brief review of population projection models

It is important to first review existing methods or models of population projection. Projections and estimates are two different terms and are commonly misinterpreted. Projection involves assumptions on future population change, and estimates relate to providing precise values of the past and present population (Rayer, 2015). Although these two terms are different, some estimation methods can also be used for projection. This section discusses projection methods and future population change in Malaysia. Since district-level units are used to project population change by settlement type, four types of methods are suitable for small-area projections: trend extrapolation, cohort-component, structural, and microsimulation models (Smith et al., 2017).

7.2.1 Trend extrapolation models

Trend extrapolation models can be used for both short-range and long-range projections and are based on historical population trends (Rayer, 2015). If historical data are not available, Armstrong (1985) proposed using a range of alternative data such as similar historical data or data from laboratory or field tests. The methods of this model vary from the simple to complex. Simple methods are linear, geometric, or exponential, while complex methods involve linear trends, polynomial curve fitting, exponential curve fitting, logistic curve fitting, and ARIMA time-series models (Smith et al., 2017). Besides total population, this model can also be used to project subgroup populations such as religious or ethnic groups or even growth components such as birth and death rates or population ratio (e.g., district population shares as part of a state's population) (Smith et al., 2017).

Earlier versions of trend extrapolation models during the late nineteenth and early twentieth centuries were not widely used and have several weaknesses (Smith et al., 2017). This type of model lacked in terms of theoretical aspect in relation to future population change and details used to examine demographic changes. However, this model has famously been used in recent years as new and more detailed methods have been developed. A well-known extrapolation models is Markov chain analysis. Markov chain analysis depends highly on the assumption that the migration matrix remains constant over the forecast period. Examples of studies that have applied Markov chain analysis for migration are Joseph (1975), Rogers (1966), and Zimmermann and Constant (2012). Recent research using the extrapolated ratio method has incorporated geographic information systems (also known as GIS) to project global population using a grid pattern (Lee, Stuart, & Robert, 2011).

Small data requirements and inexpensive, fast, and easy application make this model a suitable choice for small-area projections when data is quite limited, and a complex projection method is not an option. Despite its simplicity, this model often provides accurate projections at least for total population, and complex projections are not necessarily better. However, there are several shortcomings of trend extrapolation models. First, long-range projections of the model are prone to more error than short-range projections (Armstrong, 1985). The model is useful for short-term rather than long-term projections because there is a small chance of major change in demographics and the social and economic conditions of a region (Champion et al., 1998). Second, the model is not able to explain current or past migration flows and only estimates or predicts total migration flows. Thirds, the application of the model is limited if geographical or administrative boundaries change over time because this reduces the amount of historical data from which to extrapolate future trends, thus making the projections prone to more error (Champion et al., 1998).

7.2.2 Cohort-component models

Besides trend extrapolation model, the cohort-component model is the most well-known and commonly used for population projection. This model accounts for three components of population change: births, deaths, and migration. A common practice in this model is to divide the population into different cohorts (by age group and sex). The components are then projected separately for each cohort. This model can also be extended by subdividing the population into more sub-groups (e.g., by ethnicity or religion). To project each component, assumptions are needed, and migration appears to be the most difficult, especially for small-area projections. Wilson (2011) proposed three approaches to project migration: 1) using a base period of migration data assuming a constant change of migration in the future; 2) using a limited amount of additional information such as expert arguments and experience of a specific region and future migration targets; and 3) using projections of other variables via techniques such as regression, econometrics, and explanatory models.

The cohort-component model is a very popular projection method for many reasons (Burch, 2018, p. 138-140). First, the standard algorithm of the cohort-component model is simple since it involves basic mathematical functions. Second, different type of data, assumptions, and application techniques can be used to apply this model. Third, this model takes into account past inputs of fertility, mortality, and migration and past population dynamics (e.g., total size, growth, age-sex composition). Fourth, the projection results are convincing because future population dynamics will follow the assumptions on future inputs (e.g., fertility, mortality, and migration rates). Fifth, this model provides guidance (especially for policy makers) to control future population dynamics. Sixth, it provides a projection with detailed age-sex composition, unlike exponential projections, which are useful for sectoral planning. Finally, this model has always been used in demographic documents for population projection.

Due to the above reasons, this model is capable of providing detailed future population projections due to demographic changes. However, it has a number of shortcomings. First, the application of this model requires high costs and extensive data and effort. Lacking or incomplete data may cause major problems with the projection process. Second, the model is computationally intensive, especially for long-range projections because a process cannot be skipped as it would affect all the intervening years. Finally, the model is limited because it does not include socio-economic determinants that shape the future paths of fertility, mortality, and migration.

7.2.3 Structural models

Following the cohort-component model, a structural model is another prominent projection method for small-area projections. This model focuses on the relationship between demographic change and non-demographic change (Rayer, 2015). Non-demographic change usually involves one or more explanatory variables. Most applications of this model typically involve migration modelling for many reasons (Smith et al., 2017). First, migration rates are more unstable than fertility and mortality rates. Second, migration commonly has a greater impact than birth and death rates on population change. Finally, economic variations have a greater impact on migration than fertility and mortality rates.

There are two types of structural models: 1) economic-demographic models and 2) urban system models. Economic-demographic models typically deal with economic variables to project migration for nations, regions, states, districts, counties, and metropolitan areas. The economic variables must be based on a theoretical perspective or framework and not be simply or randomly chosen. Examples of this model include econometric models, labour supply and demand models, population or employment ratio models, and regional economic models (Smith et al., 2017). On the other hand, urban system models are commonly used to project population, economic activities, housing, and land use at a smaller scale than economic-demographic models: census, residential, or traffic zones. Aside from common economic variables (e.g., types of occupations or employment, or unemployment rates) applied in the former model, urban system models incorporate more detailed variables such as distance, land price, development, and travel costs. A downside of these models is they require an extensive range of data, more time, and high costs for implementation.

Unlike cohort-component models, structural models, economic-demographic models, and urban system models require a wide range of data, mathematical algorithms, and theoretical frameworks for implementation. These models also introduce a further set of assumptions (e.g., the future development of non-demographic factors and their relationship to demographic factors) and can incorporate a different set of tools such as GIS to display the results.

7.2.4 Microsimulation model

A microsimulation model differs from the previously described traditional demographic models; they focus on individuals or households rather than the total population. The main idea of this model is that aggregate behaviour is determined by decisions made by individuals. Hence, it is important to model individual activities to distinguish aggregate change. The models deal with individual characteristics (e.g., age, sex, marital status, or income) rather than aggregate data (e.g., total population) to avoid aggregation bias and to produce more detailed projections. To apply this type of model, deterministic or stochastic parameters are applied to each individual to model the annual process of births, deaths, and migration. The results reflect aggregate change across populations and geographic areas (Smith et al., 2017). Microsimulation models have long been used for policy analysis and for projections (e.g., spatial and non-spatial projections). However, the application of the model is extremely complex, requiring extensive resources and data.

7.2.5 Summary

Each of the different types of projection models surveyed uses different data and methods to produce a range of outputs, from simple to complex. Trend extrapolation models are still commonly used for small-area projections despite the dominance of cohort-component models and the development of structural and microsimulation models. The advantages of trend extrapolation models are low cost and less time and data requirements. However, the model focusses only on aggregate change (e.g., total population), excluding demographic or component changes, which is not useful when attempting to examine determinants of population growth. This can lead to unrealistic outcomes because the basic assumptions of this model are based solely on historical trends. Further, the observed overall trends can mask diverging trends in the underlying demographic process that may lead to change in the future overall trajectory. In contrast, cohort-component models account separately for each component of population change (births, deaths, and migration) and can incorporate different techniques, data, and assumptions for future trends. Contrary to trend extrapolation models, application of cohort-component models requires more data (e.g., they require mortality, fertility, and population data by age and sex), more extensive computations (e.g., more calculations are required for long-range projections), and more time (e.g., to clean and verify the data) and can be quite expensive. A key advantage of cohort-component models over trend extrapolation models is their outputs are more finely disaggregated, not only by components (number of births, deaths, and migration) but also by age-sex cohort.

Conventional cohort-component and trend extrapolation models are both driven by the extrapolation of past trends. However, structural and microsimulation models can be incorporated into a cohort-component framework to explain non-demographic determinants of population growth. For example, some cohort-component models incorporate structural models to project migration. Structural and microsimulation models are able to address a wide range of questions and are important for scenario and simulation analysis. Although other models can also be used for scenario and simulation analysis, structural and microsimulation models allow the examination of a wider range of explanatory variables. Their results can show the impact of economic activities such as high wages and employment on migration patterns, or the impact of pricing mechanism on people's decisions to live in certain areas or migrate to others. On the downside, these models require more data resources and are more difficult to implement, more computationally intensive and very costly. Although structural and microsimulation models are more complex than traditional models, there is no evidence that they provide more accurate results than simpler models (Smith et al., 2017).

7.3 Data and Methods

7.3.1 Selecting the best methods/ models

According to Smith et al., (2017), each of different types of projection models uses different data and methods to produce a range of outputs, from simple to complex. In order to select the best methods/models for this chapter's projection, it is important to first identify the availability and type of data. Once the data has been identified, the attention will turns towards discussing which model fits the best. Table 7.1 shows the availability of official data obtained from the Department of Statistics Malaysia.

Table 7. 1: Availability of data

Category	Data	Source
Total population	• Malaysia total population by total, sex and age group by District level (exclude under-enumeration), 2010	Department of Statistics Malaysia
	• Malaysia total population by total, sex and age group by State level (include under-enumeration), 2010	
	• Non-Malaysian citizens by total at District level (exclude under-enumeration), 2010	
	• Non-Malaysian citizens by total, sex and age group at State level (include under-enumeration), 2010	
Birth	• Total Fertility Rate and Age-Specific Fertility Rate of Malaysia total population by State and National levels, 2010	
	• Future target of Total Fertility Rate of Malaysia total population by National level, 2040	
Death	• Total Mortality Rate and Age-Specific Mortality Rate of Malaysia total population by sex and State and National levels, 2010	
	• Future target of life expectancies at birth of Malaysia total population by National level, 2010-2040	
Migration	• Migration matrix of previous residence (origin) 5 years and Current residence (destination) by total population and District level, 2005-2010.	

Based on Table 7.1 the data allows for the application of both Trend extrapolation and Cohort-component models. However, the data are more compatible with the Cohort-component models requirement which requires birth, death and migration inputs. Cohort-component model is able to provide a detailed projection than Trend extrapolation model since the model accounts each component of population change. A major advantage of Cohort-component models over Trend extrapolation models is their outputs are more finely disaggregated, not only by each population change components but also by age-sex cohort. Furthermore, different techniques, data, and assumptions for future trends can also be incorporated in the Cohort-component model than just simple extrapolation analysis which can lead to unrealistic outcomes because the basic assumptions of Trend extrapolation model is solely based on historical trends and overall change. Further, the observed overall trends in Trend extrapolation model can mask diverging trends in the underlying demographic process that may lead to change in the future overall trajectory. Thus, Cohort-component model is much a better option than Trend extrapolation model so far.

Structural and Microsimulation models on the other hand are not suitable to be applied because 1) the purpose of this chapter is to only project and observe future population and migration change instead of identifying or explaining the determinants of future change; and 2) these models require more data resources, more difficult to implement, more computationally intensive and very costly. Although the Structural and Microsimulation models can be incorporated into a Cohort-component framework, the models typically involves economic variables (e.g. employment, supply and demand, housing price, and land use) which falls beyond the scope of this thesis. Similarly, models built in Chapter 6 (unconstrained spatial interaction models in Section 6.6.5) cannot be used to predict and explain future migration flows, because they require a wider range of economic assumptions of explanatory variables on future social and economic conditions (e.g. future assumptions or targets of unemployment rate, types and status of occupations, etc.) which is a hazardous exercise. Although the Structural and Microsimulation models are more sophisticated than the Trend extrapolation and Cohort-component models, there is no evidence that they provide more accurate results (Smith et al., 2017). Due to these reasons, the option of applying Structural and/or Microsimulation models is discarded. Hence, this has left Cohort-component model as the best option for this chapter's projection method.

In order to apply the Cohort-component model, six sub-sections are drawn in order to explain the modelling approach systematically: First, a review of official projection methods and limitations in Section 7.3.2; developing a base population in Section 7.3.3; birth projections in Section 7.3.4; death projections in Section 7.3.5; internal migration projections in Section 7.3.6; and segregation of sex in Section 7.3.7.

7.3.2 Official projection review and limitations

The purpose of reviewing the official projection is to have a benchmark for this chapter projection version. There are two versions of the official projections produced by the Department of Statistics Malaysia: a first version published in 2012;

and Second version published in 2016. This section will focus on the latter version since it is an updated version from the former. In general, the official projection is done by sex, single age, and each ethnic group first by State followed by National level. The Department of Statistics Malaysia also used Cohort-component model combined with other projection methods (e.g. interpolation and extrapolation for birth projections; interpolation, extrapolation, brass logit system and Sprague multiplier for death projections; transition probability matrix for internal migration projections; and exponential growth rate and linear interpolation for international migration projections).

However, there is one major problem found in the official projection: they only partially project international migration. International in and out-migration of Malaysians citizens are ignored. The non-Malaysian citizens are divided into two categories: permanent resident and non-permanent resident. Permanent residents are non-Malaysians who settled permanently while non-permanent residents are characterized by foreign workers, foreign students, foreign visitor or tourist, expatriates and others. In terms of methods, the Cohort component model is used to project the permanent residents while the projection for non-permanent residents is based on future targets obtained from related agencies (e.g. Ministry of Home Affairs predicted that foreign workers will increase from 1.9 million to 2.3 million in 2010 and 2020 respectively).

This chapter will not account the non-Malaysian citizens in the projection due to problems mentioned in the previous paragraph (e.g. partial projection of international migration) and data limitations: 1) Existing data only provides total numbers for all non-Malaysian citizens, failing to separate them into permanent and non-permanent residents; and 2) There are no detailed inputs available to help project the permanent resident of non-Malaysian citizens (e.g. birth, death and migration rates) and non-permanent resident (e.g. data are limited and only available for future target of foreign workers and foreign students).

Nevertheless, there are also other limitations existed as shown in Table 7.2. Table 7.2 shows the comparison of available data and data that are required to perform the desired projection using the Cohort component model.

Table 7. 2: Comparison between available data and required data for modelling purpose

Model component	Required data	Available data	Corresponding section (s)
Base population	Malaysian citizens by total, sex and age group at District level (include under-enumeration), 2010	Malaysia population by total, sex and age group by District level (exclude under-enumeration), 2010	Section 7.3.3
		Malaysia population by total, sex and age group by State level (include under-enumeration), 2010	
		Non-Malaysian citizens by total at District level (exclude under-enumeration), 2010	
		Non-Malaysian citizens by total, sex and age group at State level (include under-enumeration), 2010	
Birth projection	Total Fertility Rate and Age-Specific Fertility Rate of Malaysian citizens by District level, 2010	Total Fertility Rate and Age-Specific Fertility Rate of Malaysia total population by State and National levels, 2010	Section 7.3.4 and Section 7.3.7
	Future target of Total Fertility Rate and Age-Specific Fertility Rate of Malaysian citizens by District level, 2010-2040	Future target of Total Fertility Rate of Malaysia total population by National level, 2040	
Death projection	Total Mortality Rate and Age-Specific Mortality Rate of Malaysian citizens by sex and District level, 2010	Total Mortality Rate and Age-Specific Mortality Rate of Malaysia total population by sex and State and National levels, 2010	Section 7.3.5 and Section 7.3.7
	Future target of Total Mortality Rate and Age-Specific Mortality Rate of Malaysian citizens by sex and District level, 2010-2040	Future target of life expectancies at birth of Malaysia total population by National level, 2010-2040	
Internal migration projection	Number of out-migrants and in-migrants by total, sex and age groups and District level, 2005-2010	Migration matrix of previous residence (origin) 5 years and Current residence (destination) by total population and District level, 2005-2010.	Section 7.3.6 and Section 7.3.7
	Future target of out-migrants and in-migrants by total, sex and age groups and District level, 2010-2040	None	

Based on Table 7.2, the data that are available and data that are needed are directly not compatible which each other: most data are available for larger spatial units (e.g. National and State levels) whereas small-area unit (e.g. District level) are needed in this chapter; and some information is not available (e.g. future migration target, future death rates, etc.). Hence, the existing data need to be adjusted, as explained in the corresponding section(s).

7.3.3 Developing a base population

The base population is known as the first component required for the Cohort-component model. The official projection used a base population of National and State levels in their projection. In contrast, this chapter good is to produce a District and settlement levels projection, therefore requires a District and settlement levels base population.

This gives a choice of two potential base population: Census 1 (no adjustment for under-enumeration) and Census 2 (adjusted for under-enumeration). The official projection used Census 2 which is a version of the 2010 National and State census adjusted for under-enumeration as their base population. Table 7.3 shows the comparison of existing censuses available for 2010.

Table 7. 3: General comparison of population censuses 2010

	Census 1	Census 2
Smallest scale	District	State
Total population	27,484,596	28,588,600

Ideally, Census 1 is better to be used as a base population in this chapter projection because it provides information at District level, but since Census 2 was used in the official projection, both censuses need to be taken into consideration. However, the main problem between these censuses is the total population is not similar. This is because Census 1 excludes under-enumeration while Census 2 includes under-enumeration. Furthermore, for Census 2 population data is rounded to three decimal places. Hence, due to a rounding error, population data in Census 2 might not 100 percent similar to the one used in the official projection.

In order to fit these censuses together, several adjustments need to be made through the application of Iterative Proportion Fitting (IPF) method. IPF is an iterative procedure that can be used to combine information from two or more sources (Deming & Stephan, 1940; Založnik, 2011 p.2).

The first step is to adjust each District population in Census 1 to include the under-enumerated population so that the sum of District populations equals the total population in Census 2. Table 7.4 below shows the example of adjustment of total population for *Batu Pahat District* in Census 1.

Table 7. 4: Example adjustment of District total population

Age group and State/District	Census 1		Census 2	Adjusted
	Batu Pahat District (a)	*Johor State* (b)	*Johor State* (c)	*Batu Pahat District* (a) / (b) * (c)
0 - 4	30,516	263,765	286,590	33,157
5 - 9	40,117	303,291	311,789	41,241
10 - 14	41,496	311,455	315,389	42,020
15 - 19	36,875	290,506	299,090	37,965
20 - 24	39,719	318,396	329,194	41,066
25 - 29	30,608	313,055	329,494	32,215
30 - 34	25,371	250,698	264,795	26,798
35 - 39	24,457	228,269	239,894	25,703
40 - 44	26,463	209,131	218,294	27,622
45 - 49	25,640	187,427	195,094	26,689
50 - 54	21,240	160,188	165,895	21,997
55 - 59	16,896	125,693	130,496	17,542
60 - 64	13,944	96,552	99,997	14,441
65 - 69	9,672	63,220	65,498	10,020
70 - 74	8,299	49,169	50,898	8,591
75+	10,589	59,625	60,398	10,726
Total	401,902	3,230,440	3,362,805	417,793

Note: Batu Pahat District is part of Johor State.

Adjustment above only changes the total population without affecting the age group proportion. Once all Districts are adjusted, the sum of all Districts in Census 1 should equal to total population by State and National in Census 2 (Table 7.3).

Since this chapter will only project Malaysian citizens, the next step is to remove the non-Malaysian citizens from the total population. First, it is important to know the existing information that is available for non-Malaysian citizens in each census (Table 7.5).

Table 7. 5: Existing information for Non-Malaysian citizens in each census

	Census 1	Census 2
Smallest scale	District	State
Number of Non-Malaysian citizens	1,276,259	2,324,500
Type of data	By total	By total, age and sex

In Census 1 the count of non-Malaysian citizens has not been adjusted for under-enumeration and vice versa for Census 2. Census 1 only provides the total number of non-Malaysian citizens by District, while Census 2 provides the counts of by total, age group and sex by State level. Ideally, all information at District level is needed including by age group and sex. Hence, the following tables and paragraphs will explain the process of fitting these two censuses together. The first adjustment is to include the under-enumerated non-Malaysian citizens in each District (Table 7.6).

Table 7. 6: Example adjustment of total non-Malaysian citizens for *Batu Pahat District*

State/ District	Census 1		Census 2	Adjusted
	Batu Pahat District (a)	*Johor State* (b)	*Johor State* (c)	*Batu Pahat District* (a) / (b) * (c)
Total Non-Malaysian citizens	18,511	262,352	269,423	19,010

Note: Batu Pahat District is part of Johor State.

The adjustment above assumes that the under-enumerated population is proportional to the number of non-Malaysian citizens in each District. Once all the non-Malaysian citizens for all Districts are adjusted, the sum should equal to the total non-Malaysian citizens in Census 2 (2,324,500). Next, the number of non-Malaysian citizens will be adjusted by age group (Table 7.7).

Table 7. 7: Example adjustment of non-Malaysian citizens by age group 1

Age group and State/District	Census 1 Total population *Batu Pahat District* (a)	Census 1 Total population *Johor State* (b)	Census 2 Non-Malaysian citizens *Johor State* (c)	Adjusted *Batu Pahat District* (a) / (b) * (c)
0 - 4	33,157	286,590	7,701	891
5 - 9	41,241	311,789	9,001	1,191
10 - 14	42,020	315,389	8,801	1,173
15 - 19	37,965	299,090	13,901	1,765
20 - 24	41,066	329,194	49,204	6,138
25 - 29	32,215	329,494	53,705	5,251
30 - 34	26,798	264,795	38,403	3,886
35 - 39	25,703	239,894	27,702	2,968
40 - 44	27,622	218,294	19,202	2,430
45 - 49	26,689	195,094	13,801	1,888
50 - 54	21,997	165,895	9,201	1,220
55 - 59	17,542	130,496	6,301	847
60 - 64	14,441	99,997	4,400	636
65 - 69	10,020	65,498	2,800	428
70 - 74	8,591	50,898	1,900	321
75+	10,726	60,398	3,400	604
Total	417,793	3,362,805	269,423	31,635

Note: Batu Pahat District is part of Johor State.

The total number of non-Malaysian citizens calculated in Table 7.7 (31,635) is different from the one calculated in Table 7.5 (19,010). This is because it ignores the total number of non-Malaysian in every District (Table 7.7) and is based on age group proportions within the Malaysia total population. Therefore, the values are again adjusted to get the right total number of non-Malaysian citizens (Table 7.8).

Table 7. 8: Example adjustment of non-Malaysian citizens by age group 2

Age group / District	Non Malaysian citizens in *Batu Pahat*		
	Table 7.4 (a)	Table 7.5 (b)	Final adjustment Total (a) * (b) / Total (b)
0 - 4	-	891	535
5 - 9	-	1,191	715
10 - 14	-	1,173	705
15 - 19	-	1,765	1,060
20 - 24	-	6,138	3,688
25 - 29	-	5,251	3,155
30 - 34	-	3,886	2,335
35 - 39	-	2,968	1,784
40 - 44	-	2,430	1,460
45 - 49	-	1,888	1,135
50 - 54	-	1,220	733
55 - 59	-	847	509
60 - 64	-	636	382
65 - 69	-	428	257
70 - 74	-	321	193
75+	-	604	363
Total	19,010	31,635	19,010

The total numbers of non-Malaysian citizens in Table 7.8 is now similar without affecting the age group proportions. Once the number of non-Malaysian citizens are finalized, it is easy to calculate the number of Malaysian citizens (Table 7.9).

Table 7. 9: Number of Malaysian citizens by age group

Age group / District	Batu Pahat District		
	Total population (b)	Non-Malaysian citizens (a)	Malaysian citizens (a) - (b)
0 - 4	33,157	535	32,621
5 - 9	41,241	715	40,526
10 - 14	42,020	705	41,316
15 - 19	37,965	1,060	36,904
20 - 24	41,066	3,688	37,378
25 - 29	32,215	3,155	29,060
30 - 34	26,798	2,335	24,462
35 - 39	25,703	1,784	23,919
40 - 44	27,622	1,460	26,162
45 - 49	26,689	1,135	25,554
50 - 54	21,997	733	21,264
55 - 59	17,542	509	17,033
60 - 64	14,441	382	14,060
65 - 69	10,020	257	9,763
70 - 74	8,591	193	8,398
75+	10,726	363	10,363
Total	417,793	19,010	398,783

In order to separate Malaysian citizens by sex, Table 7.10 shows the calculation of the number of Malaysian citizens by age group and sex.

Table 7. 10: Number of Malaysian citizens by age group and sex

Age group and State/District	Census 1 Total population		Census 2 Male population	*Batu Pahat District* by age and sex	
	Batu Pahat District (a)	*Johor State* (b)	*Johor State* (c)	Male (d) = (a) / (b) * (c)	Female (a) – (d)
0 - 4	32,621	278,887	144,295	16,878	15,743
5 - 9	40,526	302,898	155,894	20,858	19,668
10 - 14	41,316	306,699	158,694	21,378	19,938
15 - 19	36,904	285,323	147,494	19,077	17,827
20 - 24	37,378	279,772	144,594	19,318	18,060
25 - 29	29,060	275,178	141,795	14,974	14,086
30 - 34	24,462	226,005	118,495	12,826	11,637
35 - 39	23,919	212,067	109,596	12,361	11,558
40 - 44	26,162	199,261	101,096	13,274	12,889
45 - 49	25,554	181,483	95,296	13,419	12,136
50 - 54	21,264	156,784	80,097	10,863	10,401
55 - 59	17,033	124,288	63,598	8,716	8,317
60 - 64	14,060	95,681	49,398	7,259	6,801
65 - 69	9,763	62,788	32,399	5,038	4,725
70 - 74	8,398	49,083	24,699	4,226	4,172
75+	10,363	57,186	26,399	4,784	5,579
Total	398,783	3,093,382	1,593,839	205,247	193,536

Note: Batu Pahat District is part of Johor State.

The calculation in Table 7.10 is the final step in creating the district-level base population for this chapter projection. The next step will involve projecting births.

7.3.4 Birth projections

In order to project the number of birth, it is important to have the Total Fertility Rates (TFR) and Age-Specific Fertility Rates (ASFR). The available official TFR and ASFR are for the year 2010 at State level and 2040 at National level (The Department of Statistics Malaysia assumed that the National TFR will decline from 2.1 in 2010 to 1.7 in 2040). However, since the aim of this chapter is to project District and settlement level projections of Malaysian citizens, therefore this section explains the transformation of existing information into the one that is required.

The first step is to outline several assumptions and do some tests. The official TFR is assumed to decline linearly between 2010 and 2040. Instead of yearly, the TFR will also need to be converted from annual to five-year TFR. To do this an example calculation is as follows:

$$
\begin{aligned}
\textit{TFR 2015} &= \textit{TFR 2010}^{*} - ((\textit{TFR 2010} - \textit{TFR 2040}) / \textit{total size of year}) \\
&= 2.16 - ((2.16 - 1.7) / 6) \\
&= 2.16 - 0.077 \\
&= 2.08
\end{aligned}
$$

*Note: The following year TFR (e.g. 2020) uses the same calculation by replacing * with previous 5-year TFR (e.g. 2015)*

The results of these calculations are shown in Table 7.11:

Table 7. 11: Implied official National TFR by 5 years gap

Year	Official TFR (include non-Malaysian citizens)
2010 (base)	2.16
2015*	2.08
2020*	2.01
2025*	1.93
2030*	1.85
2035*	1.78
2040 (target)	1.70

*Note: *implied TFR*

Second, the TFR and ASFR are tested by projecting birth between 2010 and 2015 and compare the results to the official birth projection. However, data on official birth projection are not publicly available. In contrast, data on the official age 0-4 population are available. Table 7.12 shows the comparison between this chapter birth projections with the official age 0-4 population in 2015.

Table 7. 12: National birth projection and comparison to official national age 0-4 population, 2015

Age group	Female 2010 (a)	ASFR/ TFR 2010 (b)	Author's birth projection 2015 (c) = (a) * (b)*5	Official age 0-4 population 2015 (d)
15 - 19	1,299,251	0.014	87,878	-
20 - 24	1,232,953	0.059	360,802	-
25 - 29	1,130,857	0.131	741,845	-
30 - 34	899,966	0.132	594,232	-
35 - 39	843,168	0.073	308,665	-
40 - 44	814,969	0.021	86,999	-
45 - 49	733,672	0.002	8,330	-
Total	6,954,835	2.161	2,188,750	2,655,481

Based on Figure 7.12, obviously, the author's projection is not the same with the official projection because the author's projection only involves projecting births while the official projection is the final projected age 0-4 population (hence producing a difference of 466,731). As has been mentioned earlier, due to unavailability of official birth data, it is impossible to compare the birth projections, thus the only solution is to project age 0-4 population so that this chapter projection produces similar values of age 0-4 population with the official projection (2,655,481). This can be done by applying the adjustment factor (Table 7.13).

Table 7. 13: Adjusting rate and age 0-4 national projection, 2015 (Continuation from Table 7.12)

Age group	Adjustment factor (e) = Total (c) / Total (d)	Adjusted rate 2010 (f) = (b) * (e)	Age 0-4 population 2015 (c) * (d) * 5
15 - 19		0.016	106,617
20 - 24		0.071	437,739
25 - 29		0.159	900,036
30 - 34	1.213241	0.160	720,946
35 - 39		0.089	374,485
40 - 44		0.026	105,551
45 - 49		0.003	10,106
Total		2.621	2,655,481

The step in Table 7.13 will produce similar values to the official national age 0-4 population in Table 7.12. The adjusted rate is now known as Total Child Rate (TCR) instead of TFR because it is based on projected children that will be born and also survive (age 0 – 4) in 2015. Once the National rates are adjusted, the next step is to transform the State TFR into TCR by applying similar approaches. Table 7.14 shows an example transformation for Johor State ASFR and TFR.

Table 7. 14: Example transformation of TFR into TCR by State

Age group	*Johor State* 2010		
	ASFR	Adjustment factor	ATCR
	(a)	(b)	(a) * (c)
15 - 19	0.012		0.015
20 - 24	0.059		0.072
25 - 29	0.135		0.164
30 - 34	0.131	1.213241	0.159
35 - 39	0.068		0.083
40 - 44	0.018		0.022
45 - 49	0.001		0.001
TFR / TCR	2.12		2.58

From now on, the TCR and ATCR 2010 will be used to project 0-4 population in 2010-2015 for both National and State levels. To compensate for using TCR rather than TFR, the Cohort component model will not separately model deaths to 0-4 years old. For the subsequent periods (2025-2040), first, the National TCR is assumed to decline linearly by adding the difference of the average official TFR in Table 7.15.

Table 7. 15: Average official national TFR and TCR, 2010-2040

No.	Year	Official TFR	5 year period	Official average TFR	Difference of official average TFR	Author's TCR
	(a)	(b)	(a1*) + (a2*)	(c) = ((b1*) + (b2*)) / 2	(d) = (c2*) – (c3*)	2.62** – (d3)**
1	2010	2.16				
2	2015	2.08	2010-2015	2.12	-	2.62 (Table 7.13)
3	2020	2.01	2015-2020	2.05	0.08	2.54
4	2025	1.93	2020-2025	1.97	0.08	2.47
5	2030	1.85	2025-2030	1.89	0.08	2.39
6	2035	1.78	2030-2035	1.82	0.08	2.31
7	2040	1.70	2035-2040	1.74	0.08	2.24

Note:
** is for 2010-2015. For the following period (e.g. 2015-2020), starting year (2015) minus end year (2020).*

*** is for 2015-2020. For the following period (e.g. 2020-2025), starting year TCR (2.54) minus d4 (0.08) and so forth.*

To check the accuracy and assumption of the National TCR, the next step is to test projecting the age 0-4 population in 2020 (Table 7.16).

Table 7. 16: Checking the National ASCR and TCR 2015-2020

Age group	Female 2015	ASCR/ TCR 2015-2020	Age 0-4 population 2020	
			Author's	Official
	(a)	(b)	(c) = (a) * (b)	(d)
15 - 19	1,271,891	0.080	101,751	-
20 - 24	1,296,391	0.345	447,255	-
25 - 29	1,229,791	0.773	950,628	-
30 - 34	1,127,592	0.778	877,267	-
35 - 39	896,594	0.431	386,432	-
40 - 44	838,794	0.126	105,688	-
45 - 49	808,344	0.013	10,508	-
Total	6,954,835	2.546	2,879,530	2,685,218

Note: ASCR is Age-specific Child Rate

Table 7.16 shows 0-4 population are over-projected by almost 200,000 people. The excess population in 2020 is equivalent to 7.2 percent difference (as of the official 0-4 population) which means they will end up accumulating a large percentage of difference in 2040. Hence, the assumption of linear TCR needs to be rejected. Reasons for the differences are probably due to the change of ethnic composition (the official projection projects birth by single age of each ethnicity using publicly unavailable TFRs), or changes in percentage of non-Malaysians, or to differences in inter-state migration leading to different levels of exposure to State-specific TFRs.

After rigorous testing and checking, the Author decided to apply similar methods to that in Table 7.14 by implementing the adjustment factor which returns 0 projection gaps for all periods (2015-2040). In other words, the projection is constrained to produce the same number of 0-4 year old at a National level as the official projection. Table 7.17 shows the final national TCR and ASCR used in this chapter projection. For the State TCR, similar process is repeated by applying the adjustment factor in Table 7.17.

Table 7. 17: Final National TCR and ASCR, Malaysia 2010-2040

Age group		Period					
		2010-2015	2015-2020	2020-2025	2025-2030	2030-2035	2035-2040
ASCR	15 - 19	0.082	0.073	0.067	0.064	0.061	0.061
	20 - 24	0.355	0.318	0.291	0.275	0.265	0.263
	25 - 29	0.796	0.712	0.652	0.617	0.595	0.589
	30 - 34	0.801	0.717	0.657	0.621	0.599	0.593
	35 - 39	0.444	0.397	0.364	0.344	0.332	0.329
	40 - 44	0.130	0.116	0.106	0.100	0.097	0.096
	45 - 49	0.014	0.012	0.011	0.011	0.010	0.010
TCR		2.621	2.346	2.149	2.031	1.960	1.941
Adjustment factor			0.895	0.916	0.945	0.965	0.990

Once the National and State TCR and ASCR are finalized, the final step is to project District level birth projection by assuming that each District has the same ASCR as the State it belongs to.

7.3.5 Death projections

It is also necessary to have Total Mortality Rates (TMR) and Age-Specific Mortality Rates (ASMR) to project future deaths. However, the only information that are publicly available is the TMR and ASMR at State level for 2010 and life expectancy for 2010 and 2040 at National level (The National life expectancy is assumed to increase by 0.2 by each year where male will increase from 72 in 2010 to 78 in 2040, and female from 77 to 83). Therefore, this section explains the process of transforming existing information to fit with this chapter purpose, which is to project deaths at District and settlement levels.

First, the existing TMR and ASMR are tested by projecting future deaths for the first five year period (2010-2015) and compare to the officially projected deaths (Table 7.18).

Table 7. 18: Comparison between death projections, 2015

Age group	Base population 2010		ASMR 2010 (include non-Malaysian)		Projected death 2015			
					Author's		Official	
	Male (a)	Female (b)	Male (c)	Female (d)	Male (e) = (a)*(c)*5	Female (f) = (b)*(d)*5	Male (g)	Female (h)
0 – 4	1,260,252	1,194,255	1.806	1.480	11,382	8,838	2,761	-37
5 – 9	1,336,249	1,269,652	0.248	0.188	1,658	1,197	1,159	-2,439
10 – 14	1,333,649	1,273,352	0.398	0.249	2,656	1,585	1,959	1,460
15 – 19	1,340,949	1,299,251	1.101	0.332	7,382	2,158	3,458	2,860
20 – 24	1,239,153	1,232,953	1.237	0.442	7,665	2,726	3,462	3,162
25 – 29	1,153,156	1,130,857	1.228	0.542	7,083	3,062	3,164	3,265
30 – 34	943,964	899,966	1.871	0.750	8,832	3,376	3,171	3,372
35 – 39	861,367	843,168	2.522	1.004	10,862	4,231	4,273	4,374
40 – 44	809,069	814,969	3.401	1.621	13,757	6,605	7,125	6,625
45 – 49	776,470	733,672	5.180	2.779	20,109	10,193	11,676	9,277
50 – 54	666,775	639,276	7.955	4.607	26,522	14,725	17,679	12,680
55 – 59	525,380	503,881	12.071	7.128	31,710	17,960	23,484	15,284
60 – 64	402,585	396,585	18.481	10.995	37,200	21,803	31,487	26,288
65 – 69	261,990	261,090	28.902	18.704	37,861	24,417	32,592	26,792
70 – 74	195,093	204,292	45.889	33.436	44,763	34,153	128,287	135,585
75+	207,792	252,990	87.007	86.657	90,397	109,616		
Total	13,313,893	12,950,207			359,837	266,645	275,735	248,547
Grand total		26,264,100				626,482		524,281

Based on Table 7.18, the numbers of deaths are over-projected by more than 100,000 people in 2015 which is a significant gap. This is mainly because the official projection used different methods in projecting death: Yearly projection, use of the *Brass Logit System*, and the application of the *Sprague multiplier* method. Differences may also be due to the change of ethnic composition, or changes in the percentage of non-Malaysians, or differences in inter-State migration leading to different levels of exposure to state-specific ASMR in the official projection.

Therefore, these rates need to be adjusted so that they will produce a similar projected number of deaths. Table 7.19 shows how adjusted National ASMRs are calculated.

Table 7. 19: National ASMR adjustment (continuation from Table 7.18)

Age group	Adjustment factor		Adjusted ASMR 2010	
	Male	Female	Male	Female
	(i) = (g) / (e)	(j) = (h) / (f)	(i) * (c)	(j) * (d)
0-4	0.243	-0.004	0.438	-0.006
5 - 9	0.699	-2.039	0.173	-0.384
10 - 14	0.737	0.922	0.294	0.229
15 - 19	0.468	1.325	0.516	0.440
20 - 24	0.452	1.160	0.559	0.513
25 - 29	0.447	1.066	0.549	0.577
30 - 34	0.359	0.999	0.672	0.749
35 - 39	0.393	1.034	0.992	1.037
40 - 44	0.518	1.003	1.761	1.626
45 - 49	0.581	0.910	3.007	2.529
50 - 54	0.667	0.861	5.303	3.967
55 - 59	0.741	0.851	8.940	6.067
60 - 64	0.846	1.206	15.643	13.257
65 - 69	0.861	1.097	24.880	20.523
70-74	0.949	0.943	43.555	31.532
75+	0.949	0.943	82.583	81.723

The adjusted rates in Table 7.19 are then used to project numbers of death and calculate population that will survive in 2015, again comparing the results to the official projection (Table 7.20).

Table 7. 20: Example calculation of national death projection and survived for male population in 2015

Male population						
Age in 2010	Population 2010	Adjusted ASMR 2010	Age in 2015	Death 2010-2015	Survived population 2015	
					Author's	Official
	(a)	(b)		(c) = (a) * (b) *5	(a) – (c)	
0 - 4	1,260,252	0.438	5 - 9	2,761	1,257,491	1,257,491
5 - 9	1,336,249	0.173	10 - 14	1,159	1,335,090	1,335,091
10 - 14	1,333,649	0.294	15 - 19	1,959	1,331,690	1,331,691
15 - 19	1,340,949	0.516	20 - 24	3,458	1,337,491	1,337,491
20 - 24	1,239,153	0.559	25 - 29	3,462	1,235,691	1,235,691
25 - 29	1,153,156	0.549	30 - 34	3,164	1,149,992	1,149,992
30 - 34	943,964	0.672	35 - 39	3,171	940,793	940,793
35 - 39	861,367	0.992	40 - 44	4,273	857,094	857,094
40 - 44	809,069	1.761	45 - 49	7,125	801,944	801,945
45 - 49	776,470	3.007	50 - 54	11,676	764,794	764,795
50 - 54	666,775	5.303	55 - 59	17,679	649,096	649,095
55 - 59	525,380	8.940	60 - 64	23,484	501,896	501,896
60 - 64	402,585	15.643	65 - 69	31,487	371,098	371,097
65 - 69	261,990	24.880	70 - 74	32,592	229,398	229,398
70 - 74	195,093	43.555	75+	128,287	274,598	274,598
75+	207,792	82.583				

Based on Table 7.20, the author’s projection is able to replicate the official projected number of total population who survived in 2015 at National level.

The next step is to calculate the ASMR for the following periods (2015-2040). However, there are no publicly available data or assumptions made for future ASMR at National or State levels. The only data available is the official life expectancy and assumptions as shown in Table 7.21: The national life expectancy is assumed to increase by 0.2 by each year where male will increase from 72 in 2010 to 78 in 2040, and female from 77 to 83.

Table 7. 21: Official expected national life expectancy 2010-2040

Year	Official expected life expectancy	
	Male	Female
2010 (base)	72	77
2015	73	78
2020	74	79
2025	75	80
2030	76	81
2035	77	82
2040 (target)	78	83

Although there are no data or assumptions made for the future ASMR, the assumption on life expectancy can be used to calculate future ASMR. Before that, it is necessary to check whether the adjusted ASMR 2010 (in Table 7.20) produce a similar life expectancy to the official 2010 life expectancy or not (see Table 7.21). The adjusted ASMRs are converted into a life expectancy via input into an abridged life table (Newell, 1990). Table 7.22 shows the calculated life expectancy of Malaysia male population using the adjusted ASMRs, 2010-2015.

Table 7. 22: Abridge life table of Malaysia male population, 2010-2015

Age group	Adjusted Male ASMR 2010-2015	n	$_nM_x$	$_na_x$	$_nq_x$	$_np_x$	l_x	$_nd_x$	$_nL_x$	T_x	e_x
0 - 4	0.44	5	0.000438	0.3	0.002187	0.997813	100,000	219	499,234	7,714,780	77.148
5 - 9	0.17	5	0.000173	0.5	0.000867	0.999133	99,781	86	498,690	7,215,546	72.314
10 - 14	0.29	5	0.000294	0.5	0.001468	0.998532	99,695	146	498,108	6,716,856	67.374
15 - 19	0.52	5	0.000516	0.5	0.002576	0.997424	99,548	256	497,101	6,218,748	62.470
20 - 24	0.56	5	0.000559	0.5	0.002790	0.997210	99,292	277	495,768	5,721,646	57.624
25 - 29	0.55	5	0.000549	0.5	0.002740	0.997260	99,015	271	494,397	5,225,878	52.779
30 - 34	0.67	5	0.000672	0.5	0.003353	0.996647	98,744	331	492,891	4,731,481	47.917
35 - 39	0.99	5	0.000992	0.5	0.004949	0.995051	98,413	487	490,846	4,238,590	43.070
40 - 44	1.76	5	0.001761	0.5	0.008767	0.991233	97,926	859	487,482	3,747,744	38.271
45 - 49	3.01	5	0.003007	0.5	0.014925	0.985075	97,067	1,449	481,714	3,260,263	33.588
50 - 54	5.30	5	0.005303	0.5	0.026168	0.973832	95,618	2,502	471,837	2,778,549	29.059
55 - 59	8.94	5	0.008940	0.5	0.043721	0.956279	93,116	4,071	455,404	2,306,712	24.772
60 - 64	15.64	5	0.015643	0.5	0.075269	0.924731	89,045	6,702	428,470	1,851,309	20.791
65 - 69	24.88	5	0.024880	0.5	0.117116	0.882884	82,343	9,644	387,605	1,422,839	17.279
70 - 74	43.56	5	0.043555	0.5	0.196391	0.803609	72,699	14,277	327,802	1,035,234	14.240
75+	82.58		0.082583		1.000000	0.000000	58,422	58,422	707,432	707,432	12.109

The next page will explain the calculation and its interpretation (Table 7.23).

Table 7. 23: Function, interpretation and calculation of abridged life table

Function	Interpretation	Calculation
n	Size of age-group	-
${}_nM_x$	Age-specific mortality rate per 1,000 people	$\frac{ASMR}{1{,}000}$
${}_na_x$	Average proportion of the time lived in the interval x to x+n	Normally, 0.5 for low mortality country and 0.3 for high mortality country
${}_nq_x$	Probability of dying between age x and x+n	$\frac{n \times {}_nM_x}{1 + n(1 - {}_na_x) \times {}_nM_x}$
${}_np_x$	Probability of surviving between age x and x+n	$1 - {}_nq_x$
l_x	Number of persons alive at age x	$l_{x-n} \times {}_np_{x-n}$
${}_nd_x$	Number of persons dying between age x and x+n	$l_x \times {}_nq_x$
${}_nL_x$	Number of persons lived between age x and x+n	$n(l_{x+n} + {}_na_x \times {}_nd_x)$
T_x	Total number of person-years lived after age x	$T_{x+n} + {}_nL_x$
e_x	Life expectancy at age x	$\frac{T_x}{l_x}$

Source: Newell (1990)

The results in Table 7.22 show that the adjusted ASMR 2010-2015 produce a life expectancy of 77.1 for male and 79.2 for female (calculation for female ASMR is not shown since Table 7.22 already demonstrated the process for male). Obviously, this differs from the official life expectancy (e.g. in 2015, official life expectancy is 73 for male and 78 for female in Table 7.21). The likely reasons for these differences were mentioned earlier. Regardless of the cause of the differences between life expectancies, this chapter will still use the adjusted ASMR 2010-2015 since they produce a similar projected number of deaths, by age and sex, as the official projection.

Once the life expectancy of the adjusted ASMR are identified for 2015, ASMR of the following periods (2015-2040) can now be calculated by using the same assumption as in the official projection: life expectancy is assumed to increase by 0.2 by each year. Table 7.24 shows the expected life expectancy for the projected periods, 2010-2040.

Table 7. 24: Author's expected national life expectancy, Malaysia 2010-2040

Period	Author's expected life expectancy	
	Male	Female
2010-15	77.1 (Table 7.22)	79.2
2015-20	78.1	80.2
2020-25	79.1	81.2
2025-30	80.1	82.2
2030-35	81.1	83.2
2035-40	82.1	84.2

These life expectancies in Table 7.24 will be the main reference in calculating ASMR for 2015-2040. The method is to adjust base year ASMR (e.g. 2010-2015) to get future ASMR (e.g. 2015-2020) that produces the expected life expectancy (e.g. 2015-2020) in Table 7.24 by applying the adjustment factor. Figure 7.1 shows the procedure of getting future ASMR.

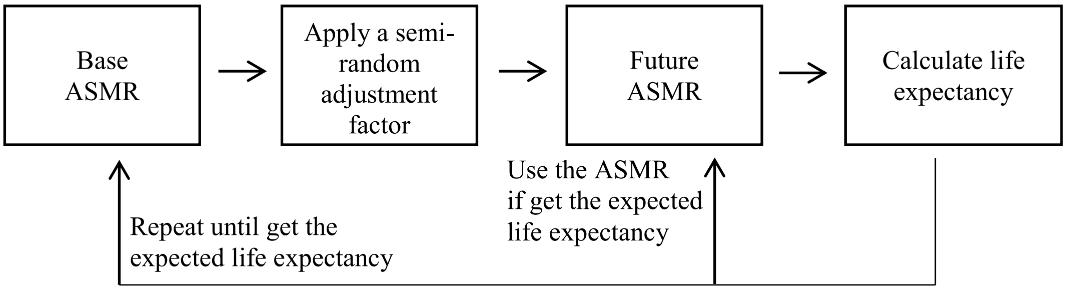


Figure 7. 1: ASMR adjustment procedure

Based on Figure 7.1, the first step is to apply a semi-random adjustment factor to base ASMR (e.g. 2010-2015) in order to get future ASMR (e.g. 2015-2020). For example, the adjustment factor should be below 1 if less death is expected or more than 1 if more death is expected based on the life expectancy pattern. The second step is to calculate life expectancy of this adjusted future ASMR. If the future ASMR produces the target life expectancy in Table 7. 24 (e.g. 78.1 for male and 80.2 for female), then it can be used for the projection. If it does not, repeat the adjustment process until getting the right ASMR that produces the exact life expectancy. Table 7.25 shows the adjustment factor that produces ASMR and life expectancy for 2015-2020.

Table 7. 25: Example calculation of Malaysia ASMR, 2015-2020

Age group	ASMR 2010-2015		Adjustment factor		ASMR 2015-2020	
	Male (a)	Female (b)	Male (c)	Female (d)	Male (a) * (c)	Female (b) * (d)
0-4	0.438	-0.006	0.940	0.940	0.412	-0.006
5-9	0.173	-0.384			0.163	-0.361
10-14	0.294	0.229			0.276	0.216
15-19	0.516	0.440			0.485	0.414
20-24	0.559	0.513			0.525	0.482
25-29	0.549	0.577			0.516	0.543
30-34	0.672	0.749			0.631	0.704
35-39	0.992	1.037			0.933	0.975
40-44	1.761	1.626			1.656	1.528
45-49	3.007	2.529			2.827	2.377
50-54	5.303	3.967			4.985	3.729
55-69	8.940	6.067			8.403	5.703
60-64	15.643	13.257			14.704	12.462
65-69	24.880	20.523			23.387	19.292
70-74	43.555	31.532			40.942	29.640
75+	82.583	81.723			77.628	76.820
Life expectancy (refer Table 7.22 for calculation)	77.1	79.2			78.1	80.2

Table 7.24 shows an example of the adjustment factors (0.940) needed to calculate ASMRs (2015-2020) that produce a life expectancy of 78.1 for male and 80.2 for female. For other periods (2020-204), the same steps are repeated as mentioned earlier. For example, to calculate ASMR 2020-2025, both male and female ASMR 2015-2020 are multiplied with 0.94. Table 7.26 shows the adjustment factor needed to calculate future national ASMR, 2015-2040.

Table 7. 26: Adjustment factor for all period

Period	Adjustment factor	
	Male	Female
2015-20	0.940	0.940
2020-25	0.940	0.940
2025-30	0.945	0.940
2030-35	0.945	0.945
2035-40	0.945	0.945

Once the National ASMRs for 2010 to 2040 are finalized, the National level adjustment factor is applied to all State ASMRs. Finally, after all State ASMR are adjusted and finalized, the final step is to project District level deaths by assuming each District has a similar ASMR to the State they belong within.

The next section will explain the final component of Cohort-Component method: migration.

7.3.6 Internal migration projections

Besides births and deaths, migration plays an important part in future population change. Migration is the fourth component of the Cohort component model. The official projection did not project international migration of Malaysian citizens due to the lack of appropriate data. For the same reason, the projection presented here also focuses only on internal migration. The main data used to project internal migration is the matrix of population by previous residence five years ago and place of current residence by District and age group, 2005-2010.

However, there is one problem identified with this dataset: out-migrants with an unknown origin are recorded in each State (Table 7.27).

Table 7. 27: Unknown out-migrant origin in each State

State	Out-migrant from unknown origin 2005-2010
Johor	31,765
Kedah	22,176
Kelantan	18,748
Melaka	14,081
Negeri Sembilan	12,492
Pahang	30,895
Pulau Pinang	32,981
Perak	27,821
Perlis	2,566
Selangor	76,017
Terengganu	15,104
Sabah	19,610
Sarawak	16,482
F.T. Kuala Lumpur	0
F.T. Labuan	772
W.P Purajaya	0
Total	321,510

These out-migrants need to be allocated an origin. Otherwise, they will be excluded from this chapter projection, leading to an underestimate of the total number of out-migrants (and, hence, in-migrants). In order to include the unknown out-migrants in the projection, the first step is to adjust the number of existing out-migrants in each District as demonstrated in Table 7.28.

Table 7. 28: Example adjustment to include unknown origin out-migrant in *Batu Pahat District*

Age group and State/ District	Total out-migrants 2005-2010			
	Known origin		Unknown origin and known origin	
	Batu Pahat District (a)	*Johor State* (b)	*Johor State* (c)	*Batu Pahat District* (a) / Total (b) * Total (c)
0-14	8,533	-	-	9,630
15-29	12,945	-	-	14,609
30-44	7,735	-	-	8,729
45-59	3,388	-	-	3,824
60 and over	1,497	-	-	1,689
Total	34,098	247,115	278,880	38,481

Note: Batu Pahat District is part of Johor State.

The calculations in Table 7.28 assumes that the age distribution of out-migrant is the same as that of migrants with a known origin. Once the number of out-migrants has been adjusted, the next step is to calculate the out-migration rate.

Since the migration data was recorded on a five-year basis (2005-2010), this section will use the middle year population (2007) to calculate the rate. However, the official censuses that provided information at District level only available for specific years: 1980, 1991, 2000 and 2010. The only information that is publicly available is the 2007 mid-year population by age-groups at State level which require an adjustment to District-level first (Table 7.29).

Table 7. 29: Example adjustment of 2007 District population

Age group and State/ District	2010 population		2007 mid-year population	
	Batu Pahat District (a)	*Johor State* (b)	*Johor State* (c)	*Batu Pahat District* based on proportion in 2010 (a) / (b) * (c)
0-14	116,418	913,768	918,123	116,843
15-29	111,246	957,778	907,367	106,250
30-44	80,123	722,983	684,913	75,935
45-59	66,227	491,485	445,464	60,039
60 and over	43,779	276,791	237,387	37,572
Total	417,793	3,362,805	3,193,253	396,639

Note: Batu Pahat District is part of Johor State.

The assumption is that the age-specific District to State population ratios observed in 2010 also applies in 2007. Given this assumption, the out-migration rate can be calculated for each District and by age groups (Table 7.30)

Table 7. 30: Example calculation of out-migration rate, 2005-2010

Age group And District	*Batu Pahat District*		
	Out-migrant 2005-2010 (a)	Mid-year population 2007 (b)	Out-migration rate 2005-2010 (a) * (b)
0-14	9,630	116,843	0.082
15-29	14,609	106,250	0.137
30-44	8,729	75,935	0.115
45-59	3,824	60,039	0.064
60 and over	1,689	37,572	0.045

Once the out-migration rates have been calculated, the next step is to calculate the in-migration rate. However, there is one problem identified if in-migration rate is used to predict future migration. Clearly, the sum of in-migrants in all Districts should equal to the sum of out-migrants in all Districts. However, the sum of in-migrants calculated using the in-migration rate is not equal to the sum of out-migrants. The main reason for this is that the number of out-migrants in each District was adjusted to allocate unknown State out-migrants (Table 7.28). By definition, the total number of in-migrants is the same as the total number of out-migrants. Hence a Districts share of all in-migrants is the same as its share of the adjusted total of all out-migrants, by age, calculated in Table 7.29. To adjust for this, the District share of all in-migrants is calculated for each age group. Table 7.31 shows an example calculation of this calculation for *Batu Pahat District*.

Table 7. 31: Example calculation of in-migrants percentage of *Batu Pahat District*, 2005-2010

Age group	2005-2010		
	In-migrants in *Batu Pahat District* (a)	Sum of in-migrants in all Districts (b)	Share of all in-migrants in *Batu Pahat District* (a) / (b)
0-14	7,356	788,248	0.0093321
15-29	8,242	1,191,067	0.0069198
30-44	6,310	712,205	0.0088598
45-59	2,978	292,343	0.0101867
60 and over	1,227	118,556	0.0103495

Source: Department of Statistics Malaysia

The next step is to test the calculated in-migration rates in Table 7.31 by predicting future in-migrants in 2015. Table 7.32 shows an example calculation of out-migrants and in-migrants for the 2015 projection in *Batu Pahat District*.

Table 7. 32: Example calculation of number of out-migrant and in-migration for 2015 projection

Age group	*Batu Pahat District* Projection 2015						
	Population (include birth and death)	Out-migration rate	Number of out-migrants	Sum of out-migrants in all Districts	Share of all in-migrants	Number of in-migrants	Sum of in-migrants in all Districts
	(a)	(b)	(a) * (b)	(c)	(d)	(c) * (d)	(e)
0-14	108,003	0.082	8,901	778,035	0.0093321	7,261	778,035
15-29	115,322	0.137	15,856	1,165,048	0.0069198	8,062	1,165,048
30-44	77,254	0.115	8,881	765,561	0.0088598	6,783	765,561
45-59	71,768	0.064	4,571	350,962	0.0101867	3,575	350,962
60 and more	51,402	0.045	2,311	170,572	0.0103495	1,765	170,572
Total	423,750		40,520	3,230,177		27,446	3,230,177

Table 7.32 shows the sum of the projected total out-migrants (c) is equal to the sum of in-migrants (e) in all Districts. Hence, this step ends the method in projecting internal migration in Malaysia. However, so far this chapter only explains the process of projecting internal migration by age group and the end result (after adding and removing the migration) is the projected total population by age group. What is required for the next year projection (e.g. 2020) is the combination of both, total population by sex and group. Therefore, the next section will explain methods on how to separate the projected population and migrants by sex.

7.3.7 Segregation of sex

For each age groups, in order to segregate the projected population and migrants by sex, this section uses the ratio of male and female in each District in 2010 and assumed it is consistent in the future. The first step is to apply this ratio to the projected number of births (Table 7.33)

Table 7. 33: Example projection of 0-4 age group by sex in 2015 at national level.

Projected Birth 2010-2015	Age group	Sex ratio		2015 projection				Difference	
				Author's		Official			
		Male	Female	Male	Female	Male	Female	Male	Female
(a)		(b)	(c)	(d) = (a)*(b)	(e) = (a)*(c)	(f)	(g)	(d) – (f)	(e) – (g)
2,655,481	0-4	0.513	0.487	1,363,572	1,291,909	1,367,690	1,287,791	-4,118	4,118

The results in Table 7.33 is a difference of only in the projected number of 4,118 males and females. As mentioned in Section 7.3.3, this is because the base population used in this chapter is slightly different from the one used in the official projection – the base population was rounded to three decimal places.

Besides birth, the sex ratio is also used after the calculation of all projection components (birth, death and migration) (Table 7.34).

Table 7. 34: Comparison of the National age-sex projection 2015.

Age group	2015 Projection*	Sex percentage		Author's projection		Official projection		Difference	
		M	F	M	F	M	F	M	F
	(a)	(b)	(c)	(d) = (a)*(b)	(e) = (a)*(c)	(f)	(g)	(d) – (f)	(e) – (g)
0 - 4	2,655,481	0.513	0.487	1,363,572	1,291,909	1,367,690	1,287,791	-4,118	4,118
5 - 9	2,451,783	0.513	0.487	1,257,307	1,194,475	1,257,491	1,194,292	-184	184
10 - 14	2,607,182	0.511	0.489	1,333,364	1,273,818	1,335,091	1,272,091	-1,727	1,727
15 - 19	2,603,582	0.508	0.492	1,322,467	1,281,114	1,331,691	1,271,891	-9,223	9,223
20 - 24	2,633,881	0.502	0.498	1,322,839	1,311,042	1,337,491	1,296,391	-14,651	14,651
25 - 29	2,465,483	0.505	0.495	1,246,182	1,219,301	1,235,691	1,229,791	10,491	-10,491
30 - 34	2,277,584	0.512	0.488	1,166,136	1,111,448	1,149,992	1,127,592	16,144	-16,144
35 - 39	1,837,387	0.506	0.494	929,543	907,844	940,793	896,594	-11,250	11,250
40 - 44	1,695,888	0.499	0.501	847,006	848,882	857,094	838,794	-10,088	10,088
45 - 49	1,610,289	0.515	0.485	829,174	781,114	801,945	808,344	27,230	-27,230
50 - 54	1,489,190	0.511	0.489	761,492	727,698	764,795	724,395	-3,303	3,303
55 - 59	1,275,691	0.511	0.489	652,231	623,460	649,095	626,596	3,136	-3,136
60 - 64	990,493	0.504	0.496	499,239	491,254	501,896	488,597	-2,658	2,658
65 - 69	741,395	0.501	0.499	371,506	369,889	371,097	370,297	408	-408
70 - 74	463,697	0.490	0.510	226,982	236,715	229,398	234,298	-2,417	2,417
75+	596,296	0.451	0.549	269,160	327,136	274,598	321,698	-5,438	5,438

*Note: * Include birth and death*
M: Male
F: Female

Result in Table 7.3.4 shows there are differences between author's projection and the official projection for all age groups. This chapter aims however is to have the same result as national when modelling age and birth at national level. Differences will inevitably arise as a result of different treatment of migration, in the distribution of ethnic groups by State and by changes over time in the sex ratio amongst age-specific deaths and migration flows. As in previous sections, an adjustment factor is applied to the sex ratio to ensure projected male and female are equal to the official projection (Table 7.33).

Table 7. 35: Adjustment of sex percentage and new projection 2015 (Continuation from Table 7.34)

	Adjustment factor		Adjusted sex percentage		Author's projection		Official projection	
Age group	Male	Female	Male	Male	Male	Female	Male	Female
	(h) = (f) / (d)	(i) = (g) / (e)	(j) = (b) * (h)	(a) * (j)	(a) * (j)	(a) * (k)		
0 - 4	1.003	0.997	0.515	1,367,690	1,367,690	1,287,791	1,367,690	1,287,791
5 - 9	1.000	1.000	0.513	1,257,491	1,257,491	1,194,292	1,257,491	1,194,292
10 - 14	1.001	0.999	0.512	1,335,091	1,335,091	1,272,091	1,335,091	1,272,091
15 - 19	1.007	0.993	0.511	1,331,691	1,331,691	1,271,891	1,331,691	1,271,891
20 - 24	1.011	0.989	0.508	1,337,491	1,337,491	1,296,391	1,337,491	1,296,391
25 - 29	0.992	1.009	0.501	1,235,691	1,235,691	1,229,791	1,235,691	1,229,791
30 - 34	0.986	1.015	0.505	1,149,992	1,149,992	1,127,592	1,149,992	1,127,592
35 - 39	1.012	0.988	0.512	940,793	940,793	896,594	940,793	896,594
40 - 44	1.012	0.988	0.505	857,094	857,094	838,794	857,094	838,794
45 - 49	0.967	1.035	0.498	801,945	801,945	808,344	801,945	808,344
50 - 54	1.004	0.995	0.514	764,795	764,795	724,395	764,795	724,395
55 - 59	0.995	1.005	0.509	649,095	649,095	626,596	649,095	626,596
60 - 64	1.005	0.995	0.507	501,896	501,896	488,597	501,896	488,597
65 - 69	0.999	1.001	0.501	371,097	371,097	370,297	371,097	370,297
70 - 74	1.011	0.990	0.495	229,398	229,398	234,298	229,398	234,298
75+	1.020	0.983	0.461	274,598	274,598	321,698	274,598	321,698

Based on Table 7.35, the 2015 projection for male and female are now similar to the official projection hence ends this section method.

In summary, all methods explained so far consider all components (base population, birth, death, and migration) required to apply the Cohort-component models. Due to different as in data availability compared to the official projections, several assumptions and adjustments are made mainly for the first five 5 year period (2010-

2015), to minimise differences that accumulate over the subsequent projected periods (2015-2040).

7.4 Comparison between author’s and official projections

Having explained how a district-level cohort-component population projection model was set up using available data sources, the attention now turns to the ensuing results. In this section, the focus is on the validation of the results. Subsequent sections go on to analyse the results and their implications for Malaysia.

7.4.1 National level

Table 7. 36: Comparison between official and author’s projections for projections for Malaysian citizens

Year	Population projection		
	Author's	Official	Difference
2015	28,395,300	28,395,300	0
2020	30,472,179	30,484,800	-12,621
2025	32,437,625	32,473,600	-35,975
2030	34,219,017	34,294,900	-75,883
2035	35,787,476	35,907,900	-120,424
2040	37,184,255	37,350,700	-166,445

Table 7.36 shows there is no difference for 2015 projections. This is to be expected because the projection components (e.g., base population, and birth and death rates) are adjusted to fit with the official projections made in 2015 (see method sections and Section 7.3 for more detail). For example, this chapter uses a similar base population, and birth and death rates are used. The main reason for this is to reduce gaps or difference between the projections made in this chapter and the official projections; hence, determining the exact values for the first five-year period is highly important. Over the following years, the differences accumulate until, by 2040, the difference is 166,455, which is equal to 0.4 percent of the official total projection. The differences in values and percentages are relatively small, indicating this chapter’s projections are more or less similar to the official projections at the

national level. The differences in the total are due to differences in assumed birth and death rates by state, and to differences in migration flows, leading to differences in exposure to state-level rates.

Besides the total population, it is important to compare the results by sex and age groups to determine which of these contribute most to the differences. Figure 7.2 compares the age-sex pyramid for the author's projections and the official projections for 2040.

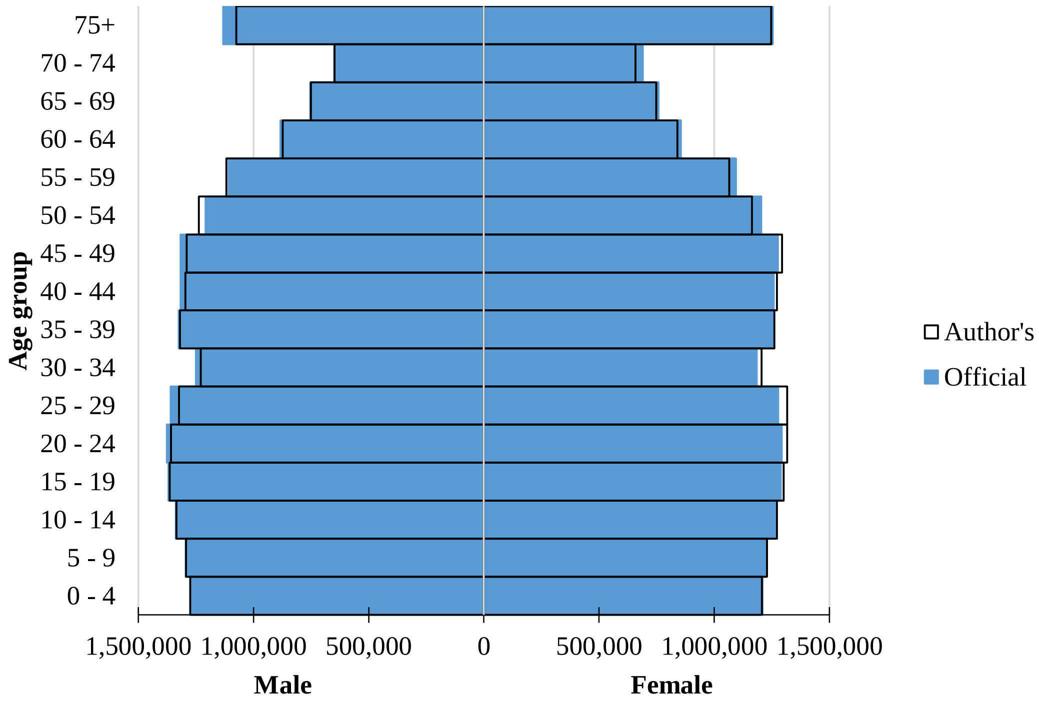


Figure 7. 2: Age-sex pyramid for the author's and official projections for Malaysia in 2040

Based on Figure 7.2, this chapter's projections replicate not only the total population but also the population by sex and age groups. The differences are small and barely noticeable. The largest difference is only 5.05 percent (males aged 75 and over) and 4.78 percent (females aged 70 to 74) (see Table 7.37). From these comparisons, the author's projections closely fit the official projections at the national level.

Table 7. 37: Differences with the official projections in percentages for Malaysia in 2040

Age group	2040 projection						Difference			Difference as of % to official projection		
	Author's			Official								
	Male	Female	Total	Male	Female	Total	Male	Female	Total	Male	Female	Total
	(a)	(b)	(c)	(d)	(e)	(f)	(g) = (a) – (d)	(h) = (b) – (e)	(i) = (c) – (f)	(g) / (d) * 100	(h) / (e) * 100	(i) / (f) * 100
0 - 4	1,274,172	1,207,415	2,481,587	1,272,193	1,209,393	2,481,587	1,979	-1,979	0	0.16	-0.16	0.00
5 - 9	1,293,407	1,228,530	2,521,937	1,294,193	1,227,993	2,522,186	-786	537	-249	-0.06	0.04	-0.01
10 - 14	1,334,579	1,272,213	2,606,792	1,337,593	1,267,493	2,605,086	-3,014	4,720	1,706	-0.23	0.37	0.07
15 - 19	1,363,404	1,301,675	2,665,079	1,368,693	1,288,793	2,657,486	-5,289	12,882	7,593	-0.39	1.00	0.29
20 - 24	1,358,766	1,317,317	2,676,083	1,375,693	1,293,693	2,669,386	-16,927	23,624	6,697	-1.23	1.83	0.25
25 - 29	1,323,206	1,317,329	2,640,535	1,359,893	1,277,393	2,637,286	-36,687	39,936	3,249	-2.70	3.13	0.12
30 - 34	1,228,988	1,205,882	2,434,870	1,249,793	1,184,594	2,434,387	-20,805	21,289	483	-1.66	1.80	0.02
35 - 39	1,319,185	1,260,600	2,579,785	1,324,293	1,258,893	2,583,186	-5,108	1,707	-3,401	-0.39	0.14	-0.13
40 - 44	1,296,267	1,271,599	2,567,866	1,317,193	1,258,193	2,575,386	-20,926	13,406	-7,520	-1.59	1.07	-0.29
45 - 49	1,290,704	1,294,422	2,585,127	1,317,093	1,276,193	2,593,286	-26,389	18,229	-8,160	-2.00	1.43	-0.31
50 - 54	1,237,205	1,163,255	2,400,461	1,208,394	1,204,293	2,412,687	28,812	-41,038	-12,226	2.38	-3.41	-0.51
55 - 59	1,117,618	1,065,733	2,183,351	1,108,594	1,093,794	2,202,388	9,024	-28,061	-19,037	0.81	-2.57	-0.86
60 - 64	873,518	840,582	1,714,100	882,895	855,195	1,738,091	-9,377	-14,614	-23,991	-1.06	-1.71	-1.38
65 - 69	750,931	748,450	1,499,381	753,296	758,696	1,511,992	-2,365	-10,246	-12,611	-0.31	-1.35	-0.83
70 - 74	647,746	657,390	1,305,136	649,697	690,396	1,340,093	-1,950	-33,007	-34,957	-0.30	-4.78	-2.61
75+	1,075,012	1,247,155	2,322,167	1,132,194	1,253,993	2,386,187	-57,182	-6,839	-64,021	-5.05	-0.55	-2.68
Total	18,784,709	18,399,546	37,184,255	18,951,700	18,399,000	37,350,700	-166,991	546	-166,445	-0.88	0.00	-0.45

7.4.2 State level

The one component of the projection model that is not adjusted to directly match the official projections is the internal migration component. Table 7.38 shows the differences between the projections of this thesis to the official projections as percentage for state level in 2015.

Table 7. 38: Differences with official projections at the state level in percentages, 2015

States	2015 projection			Difference		Difference as of % to official	
	Author's		Official				
	Without migration	With migration	With migration			Without migration	With migration
Johor	3,338,479	3,291,369	3,338,942	-463	-47,573	-0.01	-1.42
Kedah	2,032,356	2,032,125	2,022,803	9,552	9,322	0.47	0.46
Kelantan	1,720,098	1,660,067	1,725,932	-5,835	-65,865	-0.34	-3.82
Melaka	844,787	840,798	850,261	-5,474	-9,464	-0.64	-1.11
Negeri Sembilan	1,042,458	1,039,379	1,026,712	15,746	12,667	1.53	1.23
Pahang	1,545,625	1,538,263	1,531,922	13,703	6,341	0.89	0.41
Pulau Pinang	1,549,020	1,565,387	1,566,281	-17,262	-894	-1.10	-0.06
Perak	2,448,231	2,407,255	2,408,803	39,427	-1,549	1.64	-0.06
Perlis	245,185	250,611	242,500	2,685	8,111	1.11	3.34
Selangor	5,511,651	5,715,636	5,569,482	-57,832	146,153	-1.04	2.62
Terengganu	1,153,544	1,123,045	1,143,656	9,888	-20,610	0.86	-1.80
Sabah	2,592,637	2,632,038	2,644,091	-51,454	-12,053	-1.95	-0.46
Sarawak	2,556,969	2,540,281	2,544,158	12,811	-3,877	0.50	-0.15
F.T. Kuala Lumpur	1,640,490	1,564,251	1,609,186	31,305	-44,934	1.95	-2.79
F.T. Labuan	88,246	91,156	87,341	905	3,815	1.04	4.37
F.T. Putrajaya	85,525	103,638	83,229	2,296	20,410	2.76	24.52

Table 7.38 presents the results if the projections are run without modelling internal migration flows; projected state population totals are reasonably similar to those of the official projections in terms of differences by percentage. If, on the other hand, the projection includes modelled migration flows, then the differences become larger (from -4 percent to +25 percent). However, ignoring *F.T. Putrajaya*, the differences only increase to between -4 percent to +4 percent, suggesting fairly close agreement with the official state projections made in 2015. *F.T. Putrajaya* is an exception because it is a small state and the observed difference of 25 percent amounts to a

difference of only 20,410 people, accounting for a relatively small0.072 percent of the total Malaysian population in 2015.

Table 7. 39: Differences with official projections at the state level in percentages, 2040

States	2040 projection			Difference as of % to official
	Author	Official	Difference	
Johor	3,986,459	4,320,989	-334,530	-7.74
Kedah	2,710,110	2,649,423	60,686	2.29
Kelantan	2,132,170	2,673,712	-541,542	-20.25
Melaka	1,063,538	1,086,537	-23,000	-2.12
Negeri Sembilan	1,322,103	1,186,164	135,939	11.46
Pahang	2,003,603	1,980,494	23,109	1.17
Pulau Pinang	1,951,778	1,923,972	27,806	1.45
Perak	2,886,175	2,786,489	99,687	3.58
Perlis	362,625	287,802	74,823	26.00
Selangor	8,349,313	7,516,940	832,373	11.07
Terengganu	1,504,769	1,668,255	-163,486	-9.80
Sabah	3,648,763	3,869,416	-220,653	-5.70
Sarawak	3,259,710	3,363,551	-103,841	-3.09
F.T. Kuala Lumpur	1,656,071	1,814,700	-158,628	-8.74
F.T. Labuan	130,701	113,837	16,864	14.81
F.T. Putrajaya	216,367	108,419	107,948	99.57

Based on Figure 7.39, it is clear that the small errors accumulate over time for the whole 30-year period, resulting in differences ranging from -10 percent to +26 percent in 2040, if *F.T. Putrajaya* is ignored. Similar to *F.T. Putrajaya*, although *Perlis* has the second largest difference by percentage, the state is small. Accordingly, the observed difference of 26 percent amounts to a difference of only 74,823 persons (0.2 percent of the overall population).

That these differences exist is not surprising because different data and methods were applied for this chapter's projection and the official projections. First, for the official projections, a transition probability matrix was used to project internal migration between states. In contrast, for this chapter's projections, out-migration and in-migration rates at the district level were used instead and hence were more detailed. It is impossible to imitate methods used in the official projections because not provide enough information and data is provided to do. Second, this chapter used the migration rate for Malaysia's total population instead of the rate for Malaysian citizens to project the migration of Malaysian citizens because internal migration data (e.g., the migration matrix) is available only by total instead of by ethnicity. Therefore, there is no disaggregated information on Malaysian and non-Malaysian citizens. Third, the inputs (e.g., birth and death rates) of this chapter's projections were adjusted to different inputs for the official projections. For example, for the 2015 projections, the death projections of this chapter were adjusted to the official death projections that already include migration. For this reason, if this chapter's projections are run without modelling migration and the official projections are compared, the values are similar and the differences in percentages are fairly small (see Table 7.38). Fourth, there is a slight difference in the base population used because the existing census data that was obtained was rounded to three decimal points (e.g., 28,395.3), thus leading to a greater difference. Fifth, the birth and death rates at the national level were adjusted and the same adjustment was applied to all states, assuming a similar adjustment for each state. This is because information only on targets and assumptions are available at national level instead of at the state level (e.g., the national total fertility rate is expected to decline from 2.1 in 2010 to 1.7 in 2040; male life expectancy is expected to increase from 72 in 2010 to 78 in 2040; and female life expectancy is expected to increase from 78 in 2010 to 83 in 2040). Finally, this chapter's projections handle the issue of the existence of unknown out-migrants in the data.

Despite these differences, this chapter's projections match the official national projections and are fairly similar to the official state projections. Generally, there are no right or wrong projections because: 1) nobody knows what will happen in the future; and 2) both migration projections are reasonable because they a) applied recent migration trends for the projection, b) applied similar future assumptions, and c) used established demographic techniques by assuming a constant change of migration in the future. According to Armstrong (1985) and Champion et al. (1998), both projection methods are useful, especially for short-term forecasting, because there is a small chance of a major change in demographics and the social and economic conditions of a region. However, projecting migration appears to be harder than projecting births or deaths, particularly at the sub-state level, because: 1) migration rates are more unstable than fertility and mortality rates; 2) migration commonly has a larger impact than births and deaths on population change; and 3) economic variations have a greater impact on migration than fertility and mortality rates (Smith et al., 2017; Wilson, 2011). Hence, with the limitations and availability of data, projecting a constant change in migration in the future is preferable.

Nevertheless, the most significant advantage of this chapter's projection is that they involve small-area units (districts) instead of the large-area units (states) that were used or the official projections. Unlike the official projections, this chapter's projections allow for a detailed assessment of migration flows not only by district level but also by settlement type, which is important for examining future urbanisation patterns in Malaysia. This is because migration is generally an important driver of the urbanisation process and, to investigate the urbanisation sequence as hypothesized in differential urbanisation theory (from urbanisation to polarisation reversal to counterurbanisation), it is necessary to observe migration patterns for each settlement type, which the official projections do not cover.

This section ends the comparison between this chapter's and the official projections. Clearly, the projections of this chapter replicate the official projections at the national level and are reasonably similar to the official state projections. As explained in the previous paragraphs, this is unsurprising because the differences are mainly due to the different data and methods used for internal migration projections. Despite the differences, both projections follow a similar concept; for both recent migration experiences were assumed to be constant in the future. The main advantage of this chapter's projections is that they are more robust in terms of achieving the final research objective of investigating future urbanisation patterns. The following section discusses projections by settlement type.

7.5 Settlement population projections, 2015-2040

Previous studies that have applied differential urbanisation theory in developing countries have argued that such countries may not experience the same urbanisation patterns as developed countries (Gedik, 2003). There is the possibility that these countries may not experience polarisation reversal or counterurbanisation as assumed by theory due to the complex nature of the countries in terms of historical, social, economic, and cultural conditions. Furthermore, no studies have attempted to predict future urbanisation patterns to confirm the hypothesis postulated in theory (i.e., that the urbanisation sequence follows a linear pattern, starting from urbanisation and moving to polarisation reversal and, finally, to counterurbanisation). Rather, existing studies have examined only current or previous years of population change and/ or migration. Therefore, these questions lead this chapter to investigate future urbanisation patterns in Malaysia through population and migration projections. This section discusses settlement projections in two parts: 1) by total population in Section 7.5.1 and 2) by sex and age group in Section 7.5.2.

7.5.1 National-level population change

Before discussing results for population change by settlement type, it is important to provide an overview of the total change of the future population. Figure 7.3 shows the projected total number (a) and percentage of growth (b) for Malaysian citizens from 2010 to 2040.

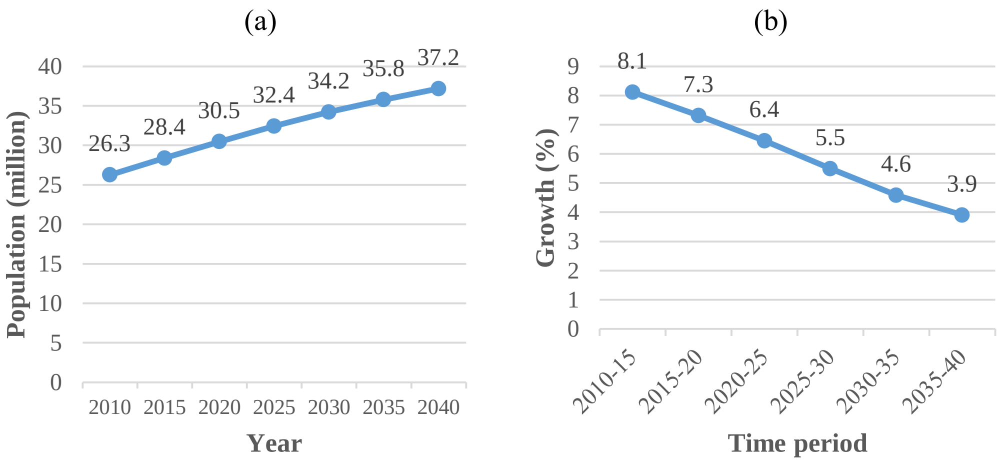


Figure 7. 3: Projected total number (a) and growth (b) for Malaysian citizens by age group, 2010-2040

Based on Figure 7.3, despite an increase in the total population, the growth rate is predicted to gradually decline over the years. Further, the age structure of the population is projected to change.

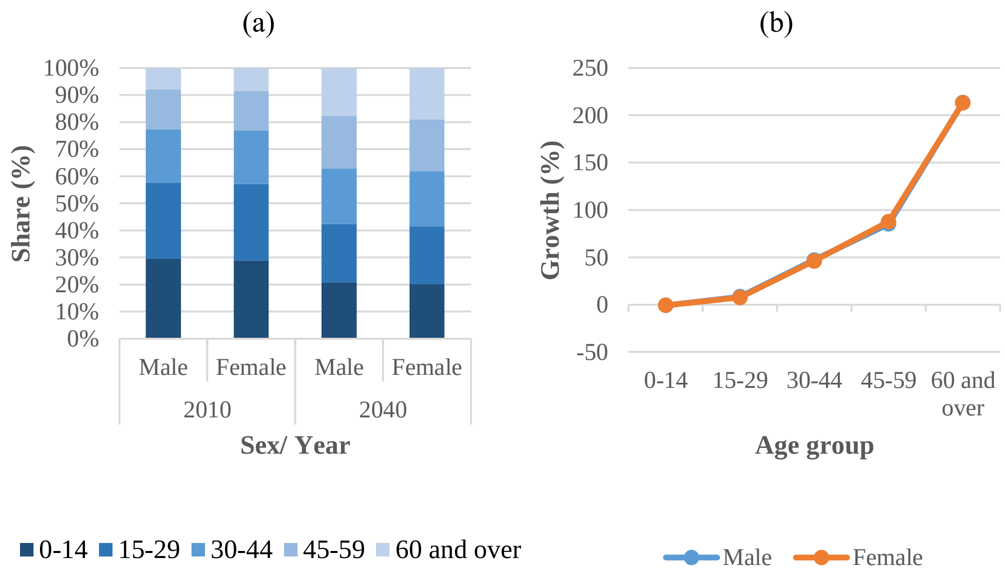


Figure 7. 4: Share (a) and growth (b) of the projected population for Malaysian citizens by age group, 2010-2040

Figure 7.4 shows that there is no significant difference between the age structure for males and females for both the starting and final years of the projections. However, the age group results display a major change in terms of pattern. Compared to 2010, the proportion of the population aged 29 and below will be less in 2040; in contrast, the proportion of the population aged 30 and above will significantly higher. This major change is explained by the exponential curve of growth between age groups, whereby the older the population, the more growth is expected. Whether this change is mainly due to fewer births or fewer deaths is explored in Section 7.6

7.5.2 Settlement level population change

Table 7. 40: Population projections by settlement type, 2010-2040

Year	Capital metropolitan			Regional metropolitan	Intermediate cities	Small towns/ villages	Remote villages
	Core	Suburban	Total				
2010	1,533,642	5,021,005	6,554,647	4,295,726	4,670,421	7,308,109	3,435,198
2015	1,564,251	5,711,922	7,276,173	4,665,797	5,029,517	7,731,926	3,691,887
2020	1,594,149	6,376,461	7,970,610	5,006,857	5,396,597	8,154,883	3,943,233
2025	1,615,709	6,973,160	8,588,869	5,316,873	5,760,367	8,582,337	4,189,179
2030	1,630,854	7,509,001	9,139,855	5,595,963	6,085,420	8,974,761	4,423,018
2035	1,643,306	7,994,605	9,637,911	5,843,391	6,362,470	9,309,370	4,634,334
2040	1,656,071	8,435,987	10,092,059	6,066,292	6,605,538	9,595,764	4,824,602

Table 7.40 indicates that total population is projected to increase in all settlement types throughout the years. By 2040, the largest population is situated in the capital metropolitan area (10,092,059) and small towns/villages (9,595,764). This is unsurprising because at start of the projections, these settlements had the highest population and were projected to continuously grow in the future. However, they have switched positions in terms of settlement size.

To check the proportion of population in these settlements, Figure 7.5 shows the percentage of the population share.

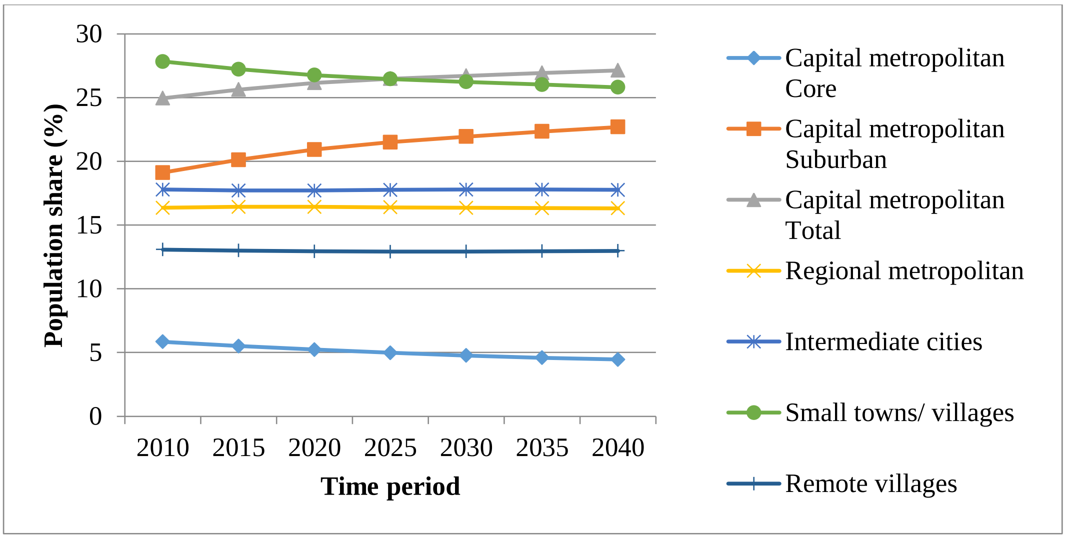


Figure 7. 5: Shares by percentage by settlement type, 2010-2040

Although the total population in all settlements is expected to increase over time, the population share results in Figure 7.5 tell a different story. Two settlement types are expected to see a significant shift in population share: the capital metropolitan area and small towns/villages. The capital metropolitan area share is expected to increase gradually, from 25 percent to 27 percent, while the share of small towns/villages is expected to decrease from 28 percent to 26 percent from 2010 to 2040. The increasing share of capital metropolitan areas will be mainly due to increases in suburban areas; the share in suburban areas is significantly higher and will gradually increase, offsetting a decline in the core city. As explained in the previous chapters, capital metropolitan suburban areas have had a higher concentration of the population since 1980 because of rapid suburbanisation. Finally, the shares of other settlements (regional metropolitan areas, intermediate-sized cities, and remote villages) will remain, with no major changes throughout the forecast period. In short, the results portray a continuation of the patterns from previous years outlined in Chapter 4.

Besides population share, another way to examine population change is by observing changes in growth by percentage (Figure 7.6 – 7.7).

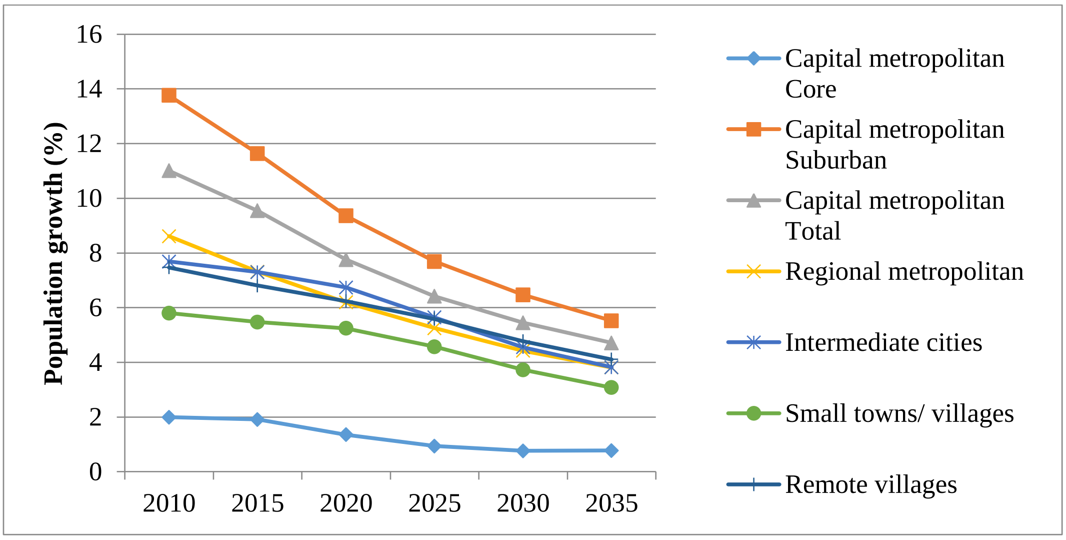


Figure 7. 6: Percentage growth by settlement type, 2010-2040

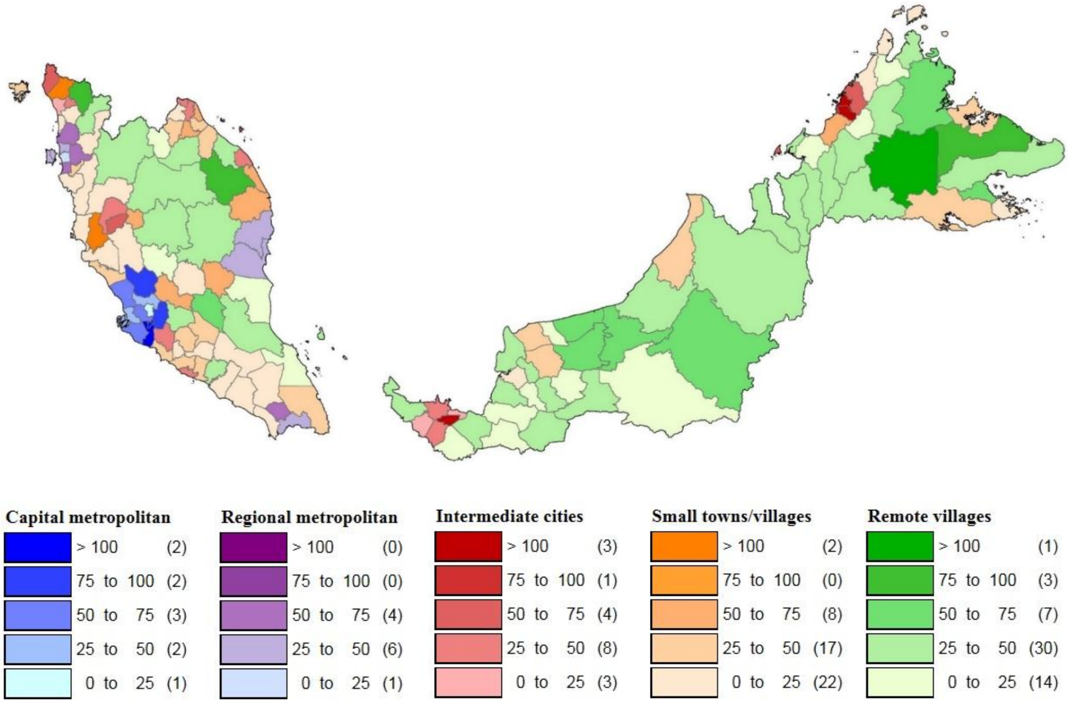


Figure 7. 7: Maps of percentage growth by settlement type, 2010-2040

Figure 7.6 indicates that all settlement types are projected to experience a continuous decline in population growth. The steepest decline in growth by percentage will be in capital metropolitan suburban areas. Over time, the growth percentage in regional metropolitan areas fall below that of intermediate-sized cities and remote villages. Despite the overall story in Malaysia being one of continuing, if not slowing, growth, Figure 7.7 shows that projected growth is more concentrated in some areas than in , particularly along the west coast of West Malaysia.

The change can also be observed through the analysis of growth relative to the initial year, 2010 (Figure 7.8). The outcome of this analysis shows which settlements are projected to change in size the most and the least over the period from 2010 to 2040.

Further, as Figure 7.8 shows, the highest change in growth is projected for capital metropolitan suburban areas, while the smallest change is projected for the capital metropolitan core.

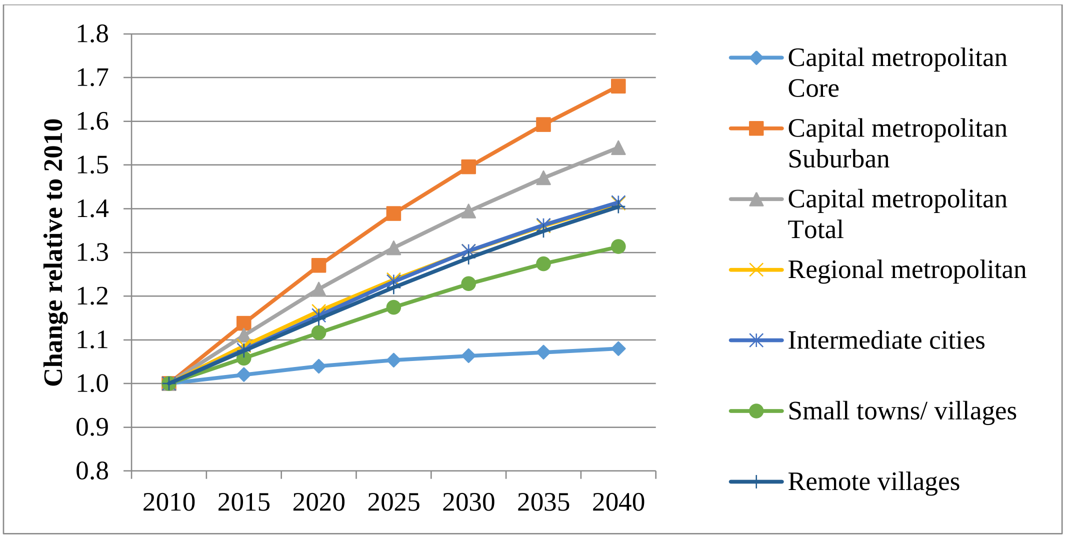


Figure 7. 8: Change relative to 2010 population by settlement type

Finally, population change can be observed relative to Malaysia as a whole (Figure 7.9). The outcome of this analysis determines which settlements will contribute the most to change in Malaysia's overall population. If the rate is above 1, this means population change will positively contribute to change in Malaysia's overall population, but if the rate is below 1, the opposite is true.

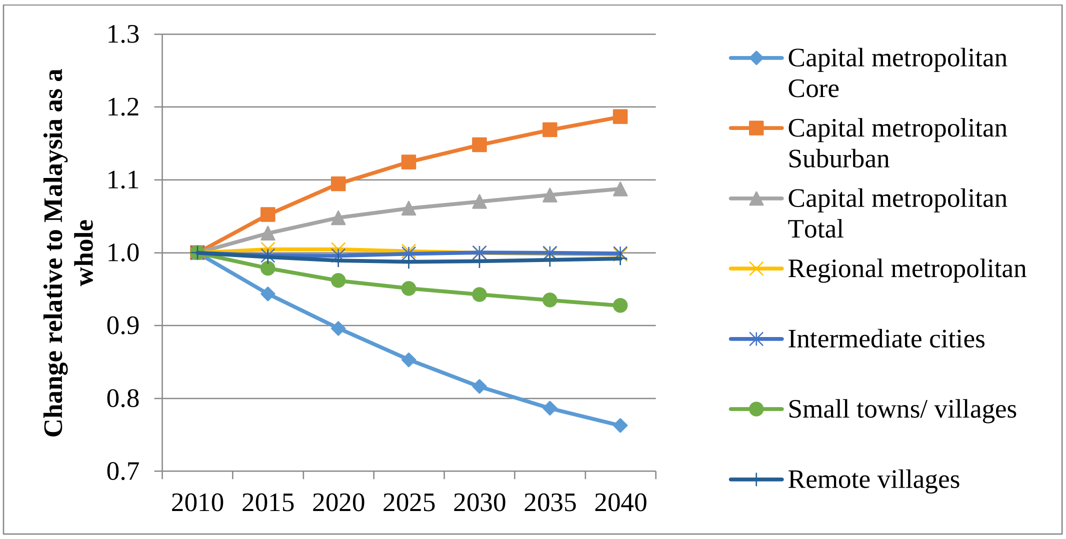


Figure 7. 9: Change relative to Malaysia's overall population by settlement type

Figure 7.9 shows the pace of growth relative to Malaysia as a whole. The results in Figure 7.9 indicate that the relative rate of population change in the capital metropolitan core is below 1 and will gradually decline throughout the projection period. Apart from the capital metropolitan core, small towns/villages also display a similar pattern. This means population change in these settlements will not positively contribute to change in Malaysia's overall population. In contrast, capital metropolitan suburban areas are projected to positively fuel change in Malaysia's overall population in the future. The population change rate in regional metropolitan areas and intermediate-sized cities, on the other hand, indicates no major changes will occur and the population will remain relatively stable over the projection period.

Overall, the population in all settlement types is projected to grow in size. However, capital metropolitan suburban areas are projected to have the most significant change in terms of population share, growth, and rate of change. This settlement type will contribute the most to change in Malaysia's overall population in the future. The primacy of the capital metropolitan core, on the other hand, is projected to diminish continuously in importance, based on not only these projections but also existing studies since 1980, as discussed in Chapter 4 and Chapter 5. Apart from the capital metropolitan core, the share and growth in small towns/villages are also projected to decline continuously in the future. Regional metropolitan areas and intermediate-sized cities, on the other hand, will see no major changes in relative standing, remaining stable throughout the projection period.

To examine population change thoroughly, the next section expands the results by examining them by sex and age group.

7.5.3 Changes in settlements by age and sex

The sex structure is expected to remain relatively stable over time. In this case, because the sex share by age will remain constant throughout the projection period, no major changes in sex shares by age will be observed (Figure 7.10).

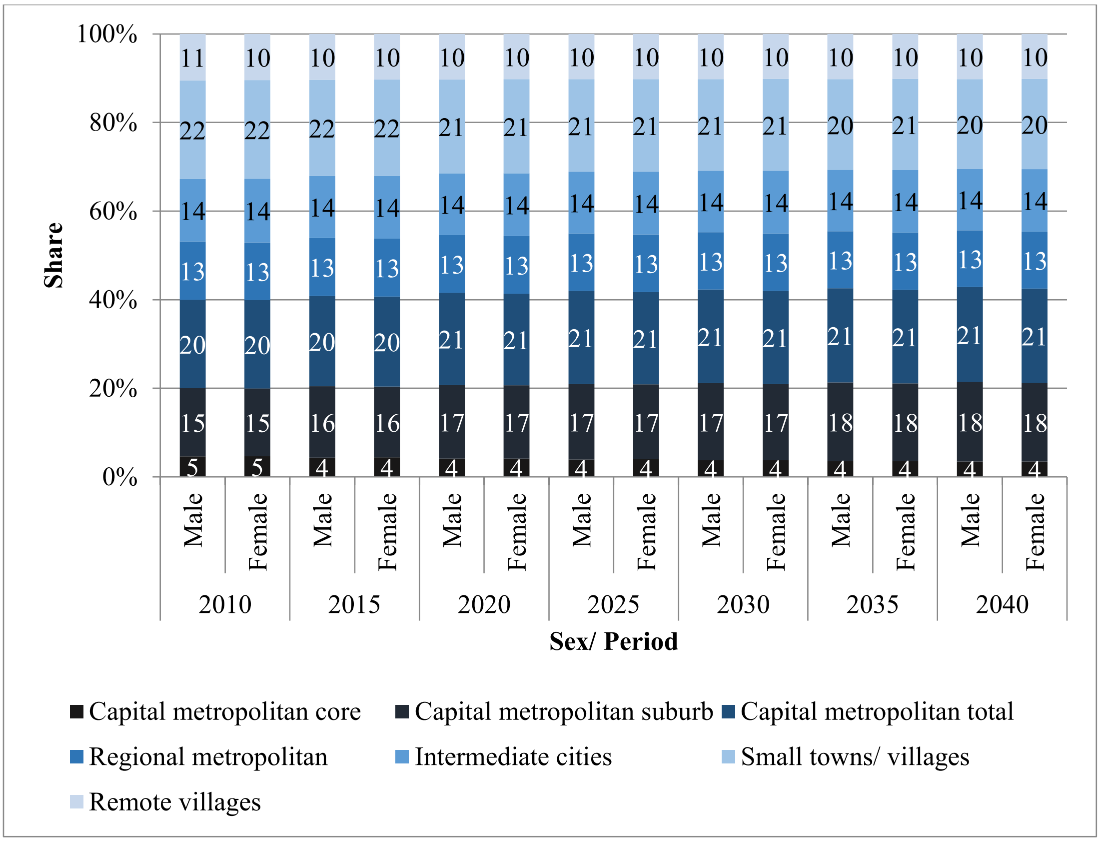


Figure 7. 10: Percentage of population share by sex and settlement type, 2010-2040

More interesting, and more dynamic, is the projected change in age structure by settlement type (Figure 7.11 and Figure 7.12). Figures 7.11 and 7.12, characterise age groups into common sets: children and adolescents aged 0 to 14; young adults aged 15 to 29, middle-aged adults aged 30 to 44, mature adults aged 45 to 59, and the elderly aged 60 and over.

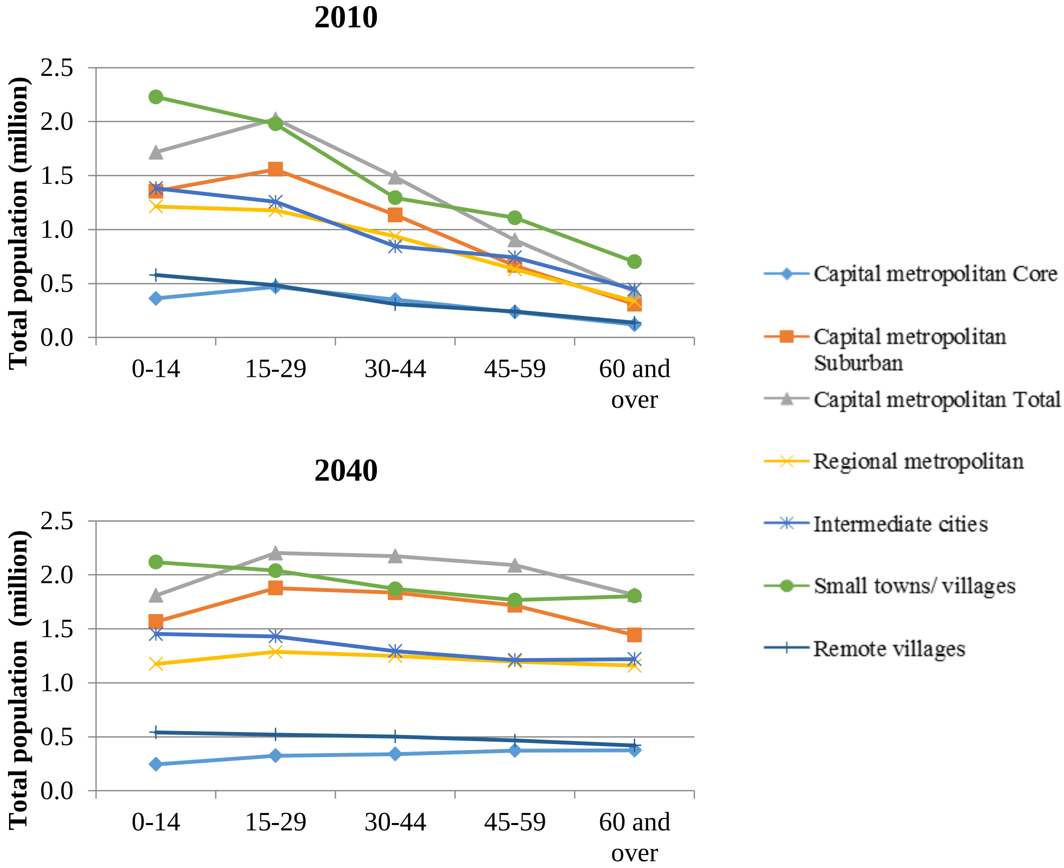


Figure 7. 11: Total population by age group and settlement type, 2010 and 2040

Clearly, the age structure in the initial and final projected years in Figure 7.11 is notably different. In 2010, younger age groups accounted for the majority of the population in most settlement types. While most settlements follow a linear age proportion (i.e., declining proportions for older groups), the capital metropolitan area (core city and suburban areas) has a larger proportion of young adults than children/adolescents. In contrast, the population in 2040 will have display a stable age proportion. Most adult groups and the elderly group (age 30 and over) are predicted to see a major increase in the future, to the point that their share of the population will be almost the same as or will exceed the population of younger age groups (children/adolescents and young adults). There will also be slight differences in the patterns of large settlements (capital metropolitan and regional metropolitan areas) and lower-level settlements. The proportion of adults and the elderly will match or exceed the proportion of children/adolescents in large settlements; in contrast, the proportion of children/adolescents will be slightly higher than the proportion of adults and the elderly in smaller settlements. To examine this further,

Figure 7.12 displays growth (a) and change (b) relative to the initial year for each age group and settlement type.

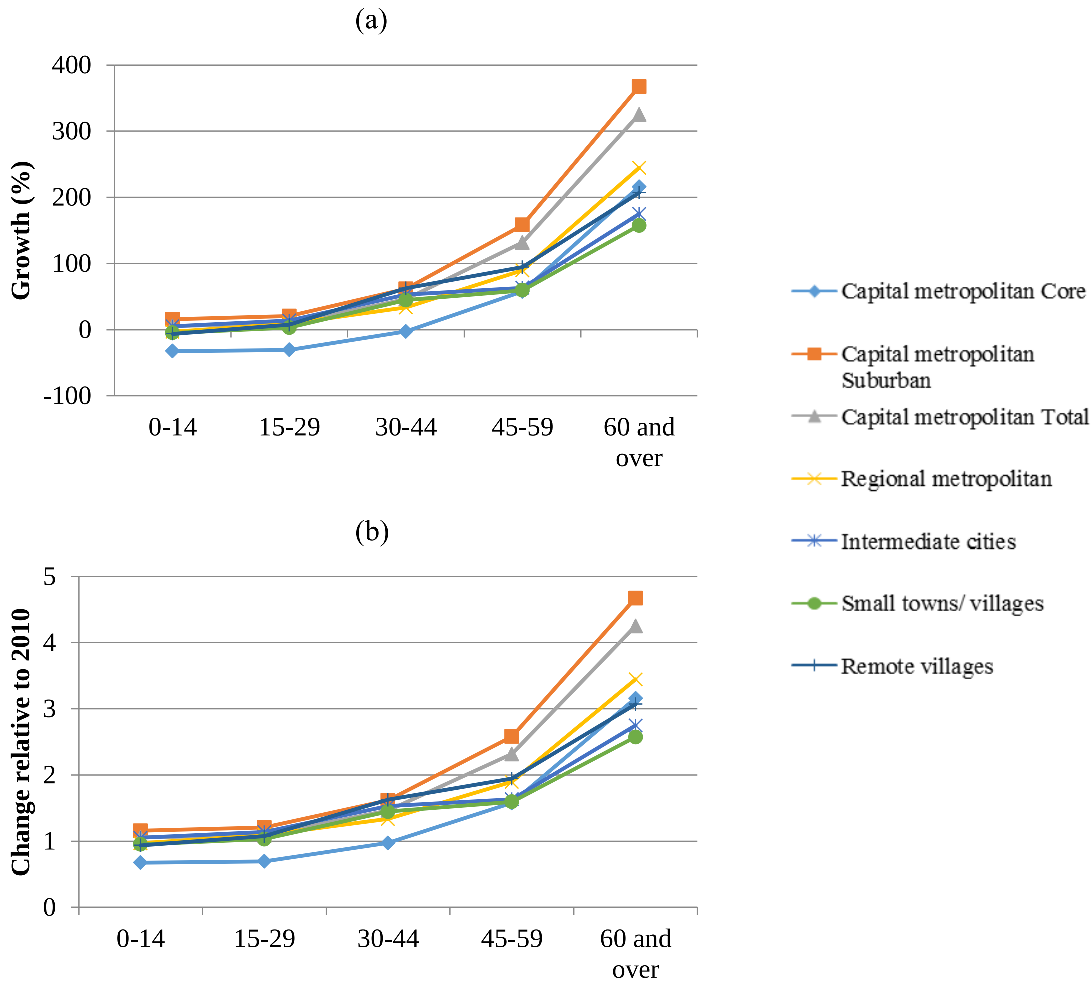
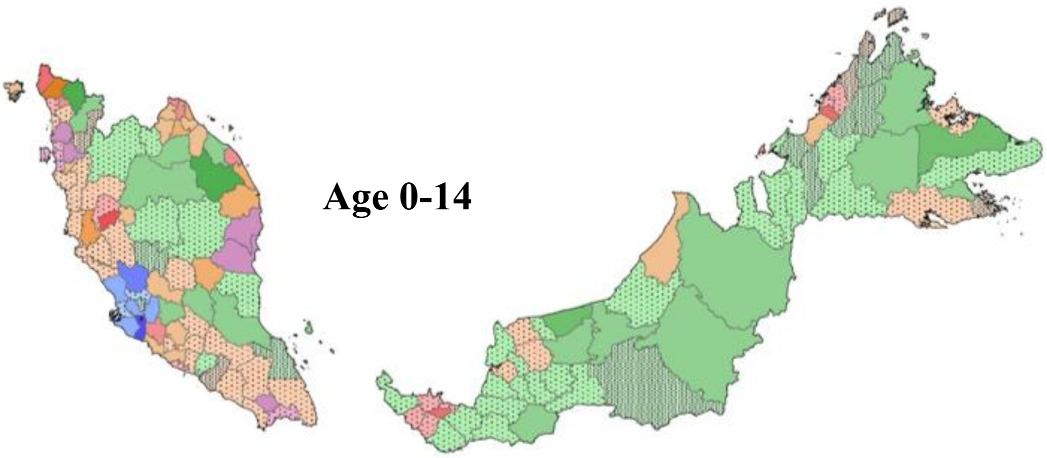
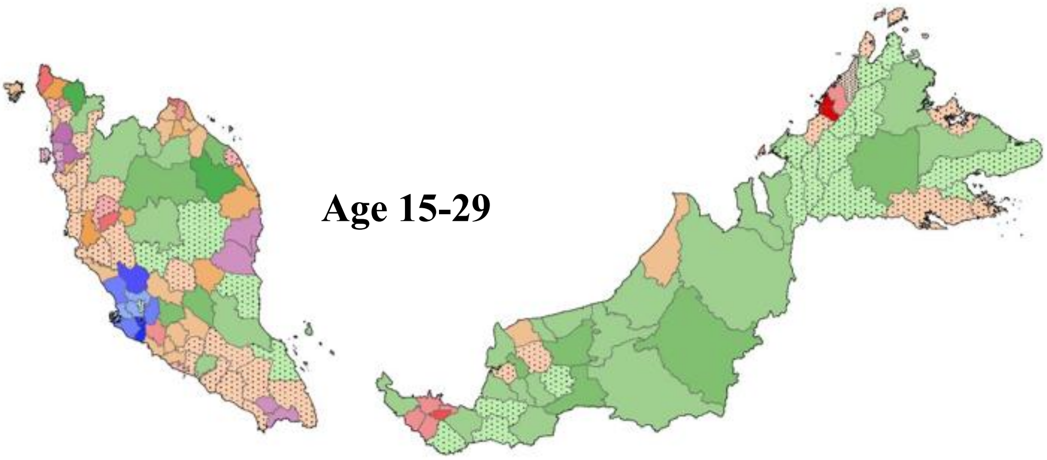
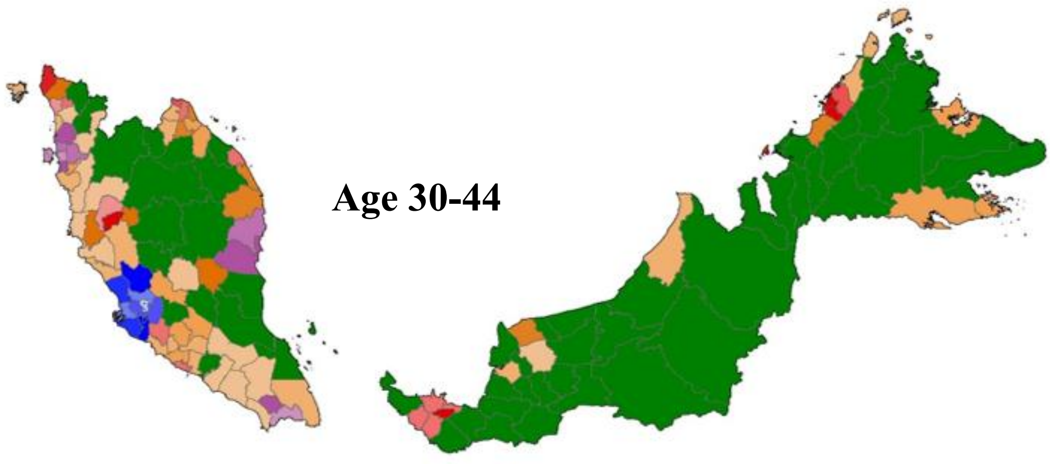


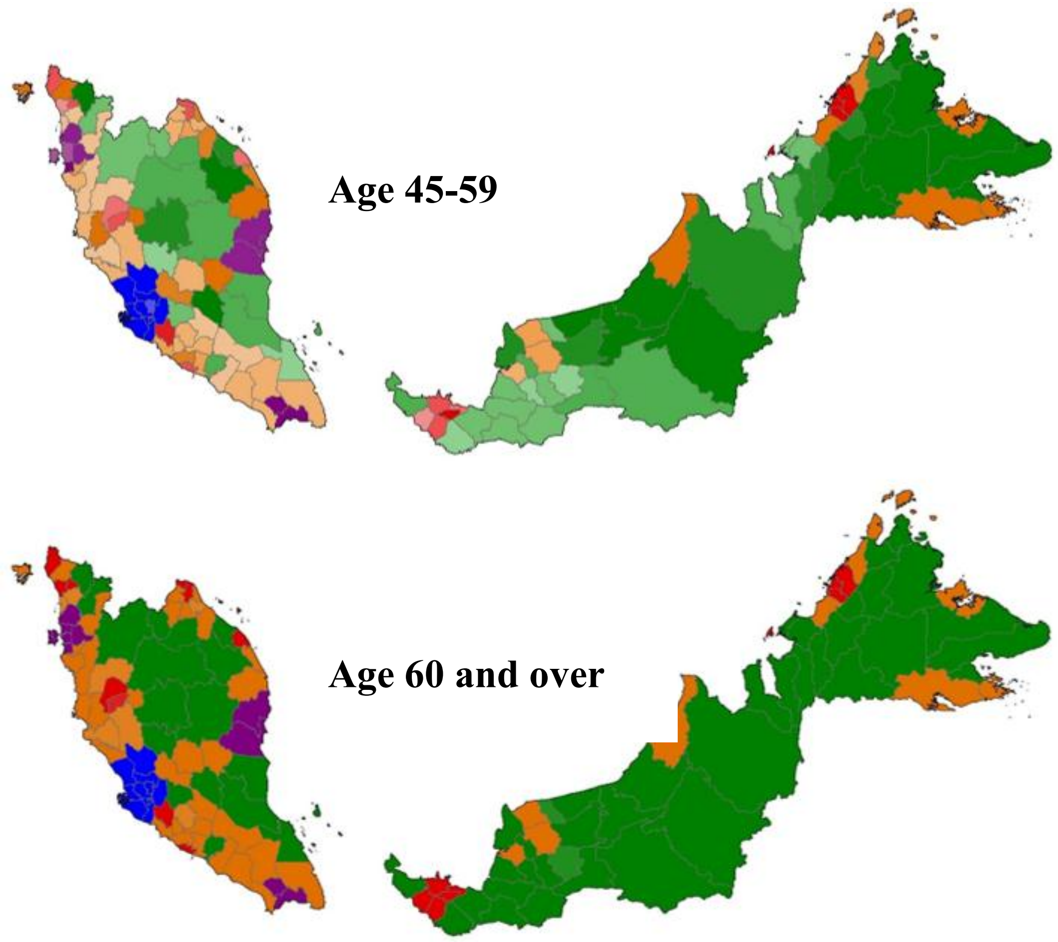
Figure 7. 12: Population growth for 2010-2040 (a) and relative change rate to 2010 (b) by age group and settlement type

The results for growth (a) and relative change (b) in Figure 7.12 have similar patterns; the growth and change rate will significantly increase as age increases, and large settlements will have the highest growth and change rates. Other noticeable results are the negative growth and change rate below 1 for children/adolescents and young adults in the capital metropolitan core. In other words, these age groups are predicted to experience a decline in total population in the future. To observe the spatial distributions of the result, Figure 7.13 illustrates the growth percentages for each age group and settlement type.


Age 0-14


Age 15-29


Age 30-44



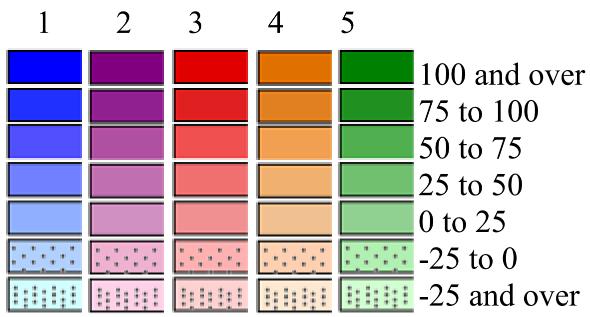


Figure 7. 13: Maps of percentage growth by age group and settlement type, 2010-2040

Note: 1 is the capital metropolitan area, 1(a) is the capital core, 1(b) is capital suburban areas, 2 is regional metropolitan areas, 3 is intermediate-sized cities, 4 is small towns/villages, and 5 is remote villages.

In summary, unlike sex, projected changes in the age structure between the starting and final years indicate that future population growth will mainly be sustained by adults (particularly adults age 30 and over) or the working-age populations (ages 15 to 65). The major increase in the adult or working-age population in the future may impact future economic growth due to an increase in the number of labourers and a shrinking proportion of children to support. This phenomenon is known as the demographic dividend. Demographic dividend occurs during the course of demographic transition when economic growth is accelerated by a decline of fertility and mortality and changes in the age structure (Gribble & Bremner, 2012). As time passes, the proportion of children grows smaller than the working-age population, consequently leading to more opportunities for a country's economic growth through the implementation of policies and investments (Gribble & Bremner, 2012).

7.6 Settlement projections: Natural increase in 2015-2040

Although it is clear that the population in different settlement types will increase by varying amounts, the cause of these changes is so far unclear. Birth and death projections play an important role in population change. The difference between births and deaths is known as natural increase. This section first discusses birth projections (Section 7.6.1), followed by death projections (Section 7.6.2), and, finally, natural increase (Section 7.6.3) for each settlement type.

7.6.1 Birth projections

Table 7. 41: Projected number of children born and surviving from age 0 to 4 by settlement type, 2015-2040

Year	Capital metropolitan			Regional metropolitan	Intermediate cities	Small towns/ villages	Remote villages
	Core	Suburban	Total				
2015	26,404	107,763	134,166	88,125	98,154	142,716	67,935
2020	24,383	111,406	135,789	84,474	102,054	148,988	69,012
2025	21,102	107,204	128,306	81,281	104,810	152,424	70,178
2030	18,325	103,899	122,224	79,251	102,760	150,165	70,248
2035	16,661	103,249	119,910	77,372	98,262	143,351	68,741
2040	15,984	104,683	120,667	76,982	95,827	137,934	67,551

Table 7.41 shows that the number of children who will be born and survive is projected to decline in all settlement types during the projection period. The change in the number of these children is based on the total female population of reproductive age (ages 15 to 49) (see Section 7.5), where the more women of reproductive age, the large the population of children. Small towns/villages will have the largest population of children, followed by capital metropolitan areas.

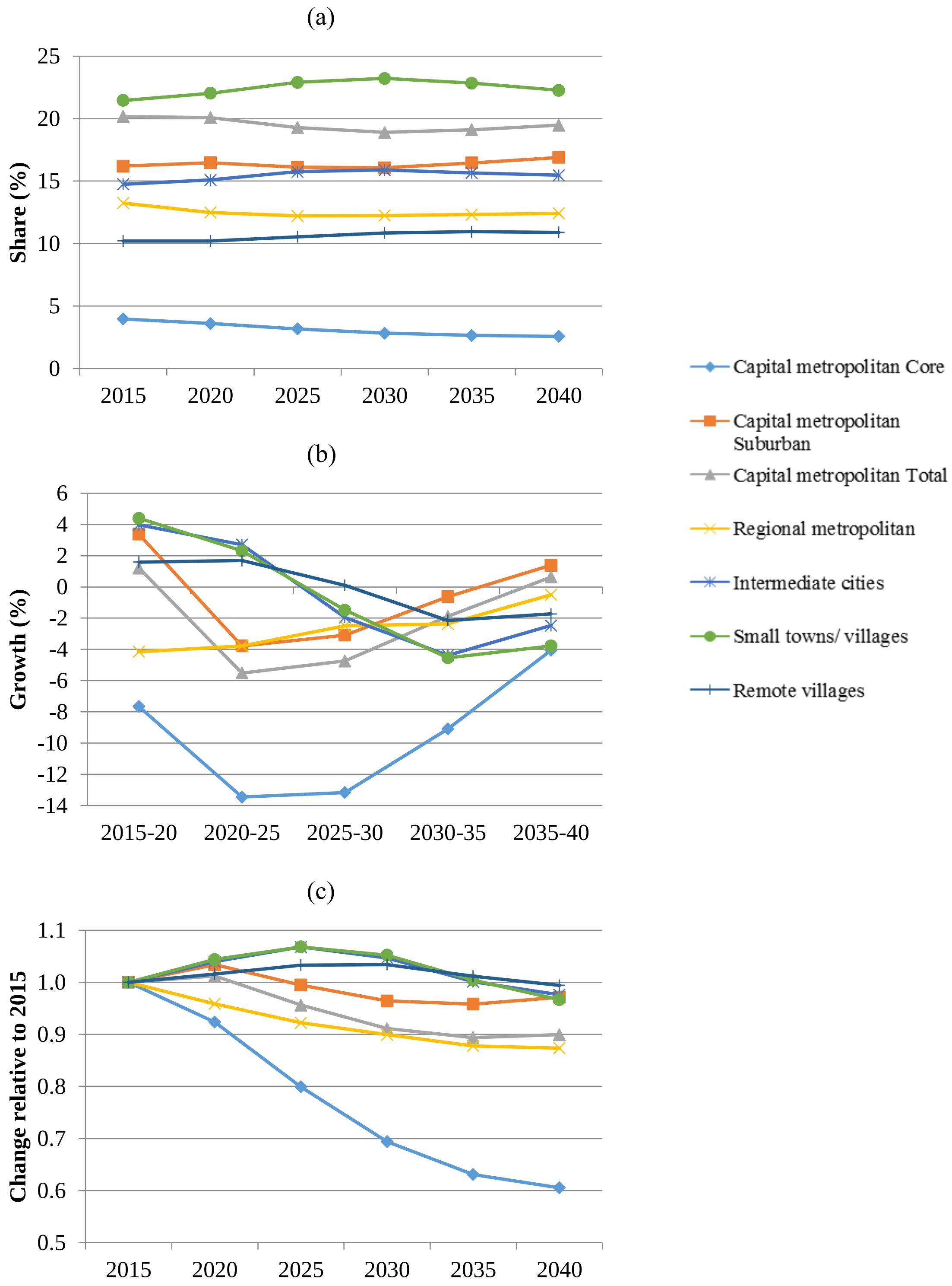


Figure 7. 14: Percentages of shares (a), growth (b), and change relative to 2015 (c) for the projected number of children born and surviving from age 0 to 4 by settlement type, 2010-2040

In terms of birth shares in Figure 7.14 (a), small towns/villages have the largest proportion because they have the largest number of births (see Table 7.41). However, few important changes can be observed (e.g., the share of births slowly declines for some settlement types while it fluctuates for others). In contrast, the growth results in (b) shows a clearer pattern; most settlements experience a decline, or negative growth, until 2025, which then diminishes in the subsequent years. In terms of relative change to the initial year in (c), the capital metropolitan core is projected to experience the fastest and largest decline of births compared to the initial year. The same situation applies for regional metropolitan areas and capital metropolitan suburban areas but at a slower rate. For other settlement types (intermediate-sized cities, small towns villages, and remote villages), births increase at a similar rate until 2025but then decline in the subsequent years. In other words, large settlements are expected to experience fewer births (especially in the capital metropolitan core) in the future while smaller settlements will follow the same trends after 2025.

The situation reflects the assumption made for child fertility rate used in the projection: child fertility rates are assumed to decline continuously through 2040. The decrease of the fertility rate in Malaysia began in the 1960s and has continued to decline to the present day. Other studies have also captured a similar result: before the 1970s, fertility trends in economically advanced developing countries in East Asia and Latin America showed a stable pattern but then declined after the 1970s (Wilson, 2011). Furthermore, although fertility rates are projected to fall below the replacement level by 2040, the total population will continue to increase in all settlement types (see Section 7.5), indicating a population momentum effect. Population momentum occurs when previous high fertility results in a large proportion of female population being of reproductive age, hence leading to a high birth rate (Keyfitz, 1971).

The contrast between patterns in large and small settlements (including rural areas) indicates that those who live in large settlements may have lower fertility than those who live in small settlements. Generally, this is because of the improvement of social and economic conditions for women (e.g., continuous usage of availability of modern contraceptives, postponement of childbearing and marriage, increasing abortion rates, higher levels of education and employment among women) (Ernestina, 2002; Hirschman, 1980). A recent fertility study in developed countries (e.g., Denmark, Finland, Norway, and Sweden) also captured similar results; the larger the settlement, the lower the fertility rate (Kulu, 2013). One of the main reasons for this is the high tendency to have more children in smaller settlements compared to larger settlements (Kulu et al., 2009). Furthermore, the location and type of housing play an important role in fertility changes; the relocation of residences to live in a bigger house or a 'family-friendly environment' evidently raises fertility levels (Kulu, 2005; Kulu & Vikat, 2007). Further, higher fertility is more prevalent among migrants than non-migrants, but has little impact due to migrant's small population share (Kulu & Boyle, 2009; Kurek et al., 2015). Fewer births in the core city may also relate to the small proportion of the female population of reproductive age, caused by large out-migration of this group to other settlement types (see Section 7.7).

7.6.2 Death projections

Following birth projections, this section explains death projections by settlement type. Table 7.42 and Figure 7.15 show the size, percentage share, and change in the growth rate relative to 2015 for projected deaths by settlement type for 2015-2040.

Table 7. 42: Projected number of deaths by settlement type, 2015-2040

Year	Capital metropolitan			Regional metropolitan	Intermediate cities	Small towns/ villages	Remote villages
	Core	Suburban	Total				
2015	5,034	14,082	19,115	17,300	22,017	33,777	12,646
2020	5,949	17,743	23,692	20,091	24,992	39,337	14,595
2025	6,987	22,175	29,162	23,389	28,417	44,102	16,827
2030	8,279	27,877	36,156	27,799	32,854	50,306	19,533
2035	9,596	33,935	43,531	32,476	37,411	56,658	22,486
2040	10,892	40,404	51,296	37,281	41,785	62,639	25,550

Table 7.42 shows the projected number of deaths in all settlement types will gradually increase as time passes. Projected deaths are based on the total population in each settlement; the larger the population in a settlement, the higher the number of expected deaths. In this case, small towns/villages have the largest projected number of deaths.

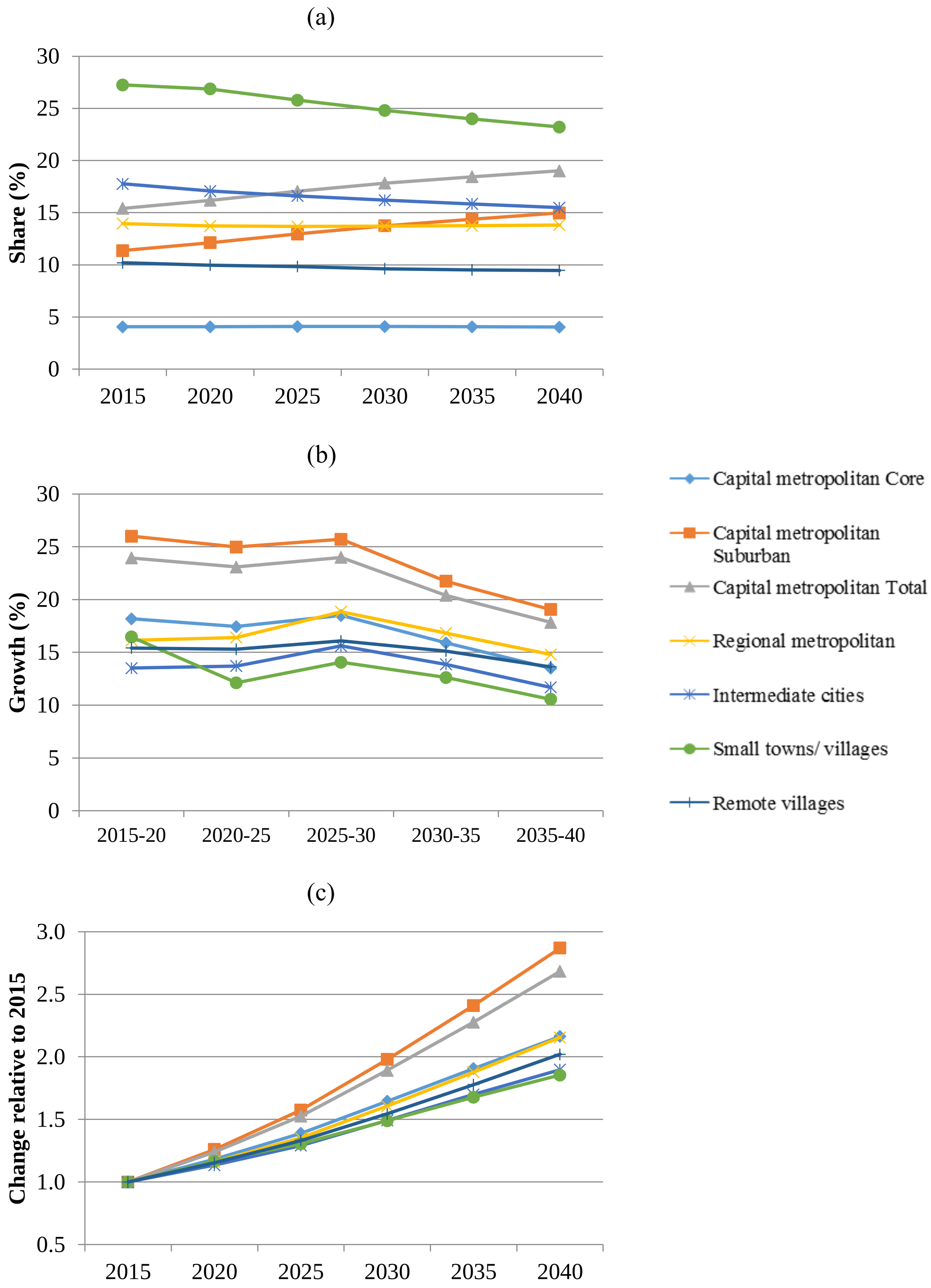


Figure 7. 15: Percentages of shares (a), growth (b) and change rate relative to 2015 (c) of projected deaths by settlement type, 2010-2040

Although the number of people dying gradually increases over the years, the results in Figure 7.15 for shares (a), growth (b) and relative change (c) tell a different story. As time passes, the share of people dying in large settlements (especially capital metropolitan suburban areas) slowly increases but decreases in smaller settlements (intermediate-sized cities, small towns/villages, and remote villages). In other words, more people are projected to die in large towns than in smaller towns. The growth and relative change results can be interpreted similarly: large towns have higher growth and change rates than smaller settlements and rural areas.

The results reflect the assumption made about future life expectancy in the projections, which is based on historical trends: future life expectancy for males and females is expected to increase continuously in the future. After the Second World War ended, life expectancy in Malaysia rapidly increased and mortality significantly declined in since at least the 1950s (Hirschman, 1980). The results also relate to work by Wilson (2011) on demographic convergence, which compared mortality and life expectancy trends in developed and developing countries for 1950-2010. The growing linear trend of life expectancy and decrease of mortality in developing countries is primarily due to growing access to health facilities and curative medicine, improvement of nutrition, and the establishment of preventive health programs (Wilson, 2011). This is supported by a recent study that found the increase of life expectancy and decrease of mortality in Malaysia was mainly due to the availability of more health care and higher socio-economic status (Chan & Kamala Devi, 2015).

Furthermore, the difference in mortality between urban and rural areas has long been debated by historians and demographers in terms of existence, causes, historical aspects, and the measurements used to examine the difference (Woods, 2003). Existing studies show that different countries display different patterns of mortality. For example, non-metropolitan areas or rural areas in the United States tended to have higher mortality than metropolitan areas during 1999-2014 (Moy et al., 2017). In contrast, due to rapid urbanisation, urban areas in European countries during the eighteenth and nineteenth centuries had higher mortality than the countryside (Woods, 2003). More detailed studies in the UK have found that while most causes of death differ between urban and rural areas, some causes (e.g., cancer, circulatory disease) are similar (Gartner et al., 2008).

Based on the above reviews, the mortality patterns in urban and rural areas in European countries and the UK are generally similar to mortality patterns in Malaysia. More deaths are projected in metropolitan areas than in smaller settlements and rural areas in Malaysia throughout the projection period. Because this chapter examines only the total number of deaths, the causes of death are explained based on theoretical findings. According to Malaysia Department of Statistics (2018), the main cause of death in urban areas in Malaysia is ischaemic heart disease, which mostly affects people aged 41 and over; in contrast, traffic accidents are the main cause of death those who aged 40 and below. Living in big cities, especially in the capital metropolitan area (Kuala Lumpur and its suburbs) is quite stressful (e.g., high living cost, high traffic, high crime) (Free Malaysia Today, 2017) and is less healthy, which can lead to heart-related diseases (e.g., respiratory diseases and lung cancer) (O'Reilly, O'Reilly, Rosato, & Connolly, 2007). Furthermore, the capital metropolitan area is known to have a record number of fatal traffic accidents since it is the busiest city in Malaysia. Therefore, it is not a unsurprising that the city has a large number of deaths (Kunasekaran, 2017).

7.6.3 Natural increase

Natural increase or decrease is interpreted as the difference between births and deaths. Results for the projected natural increase are shown in Table 7.43 and Figure 7.16.

Table 7. 43: Rate of natural increase (per 1,000) by settlement type, 2015-2040

Year	Capital metropolitan			Regional metropolitan	Intermediate cities	Small towns/ villages	Remote villages
	Core	Suburban	Total				
2015	14	16	16	15	15	14	15
2020	12	15	14	13	14	13	14
2025	9	12	12	11	13	13	13
2030	6	10	9	9	11	11	11
2035	4	9	8	8	10	9	10
2040	3	8	7	7	8	8	9

There are two important findings in Table 7.43. First, the positive values indicate that there will be more births and surviving children than deaths, hence there will be a natural increase instead of a natural decrease. Second, the natural increase rate is expected to decline in all settlement types during the projection period of 2015-2040. To examine this further, results for shares, growth, and change relative to 2015 for natural increase are shown in Figure 7.16.

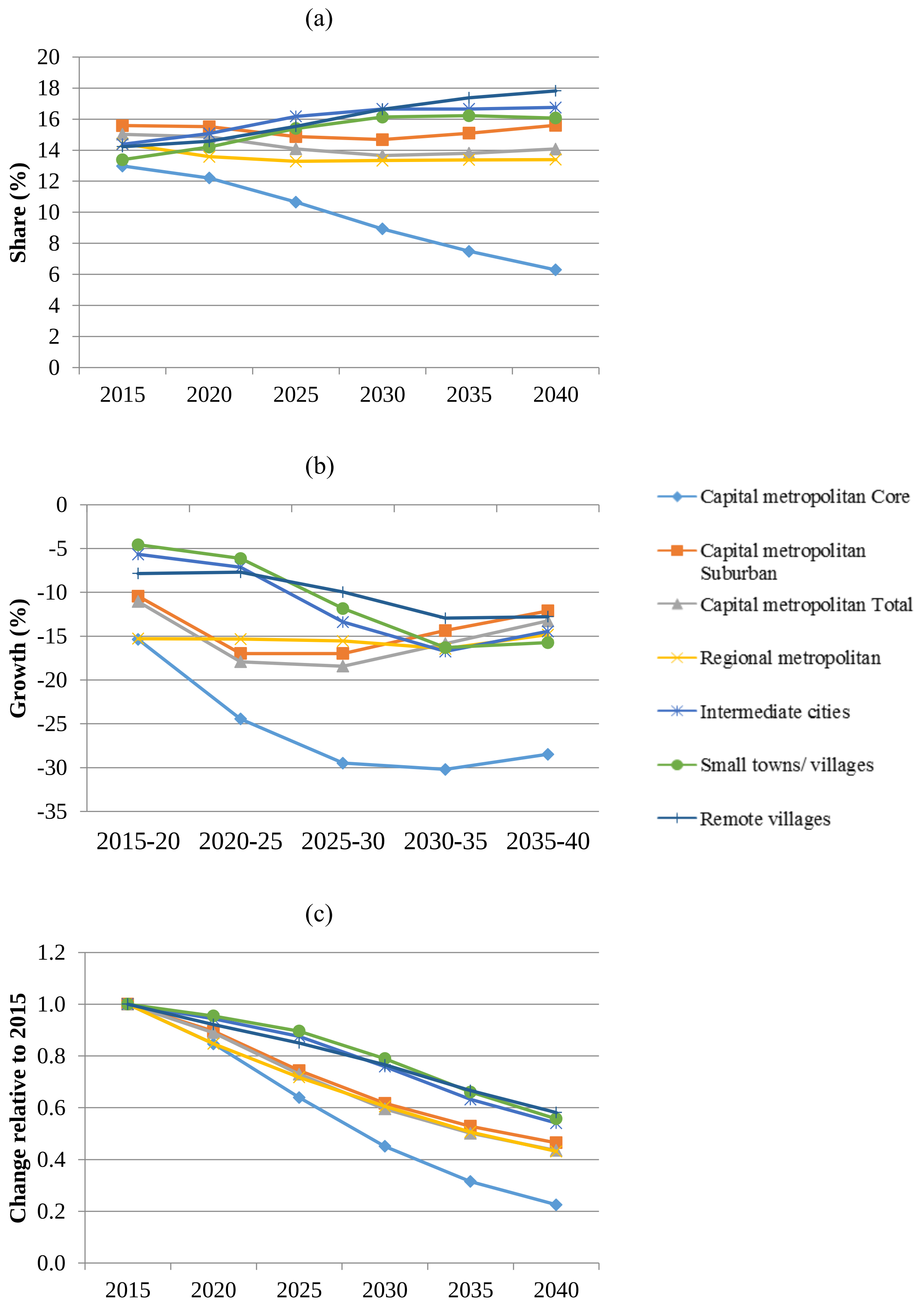


Figure 7. 16: Percentage of shares (a), growth (b), and change (c) relative to 2015 for natural increase by settlement type, 2010-2040

The share results in Figure 7.16 (a) show that natural increase is expected to decline in large settlements (especially in the core city) except in capital metropolitan suburban areas. In contrast, the share in smaller settlements will gradually increase. In other words, the number of people dying will catch up with the number of children born and surviving in large settlements while the opposite is true for smaller settlements. In terms of growth (b), all settlement types will display negative growth, with the largest change in the core city. Furthermore, the results for relative change (c) show large settlements (the capital metropolitan core and suburban and regional metropolitan areas) appear to have higher change rates than smaller settlements.

Theoretically, the change in natural increase is highly associated with the demographic transition theory. Reher (2004) reviewed the theory from a global perspective by examining mortality and fertility trends between countries. The results from the study confirm the validity of the theory. All countries experienced similar mortality and fertility trends: mortality decline preceded fertility decline. This is also true in the Malaysian context. After the Second World War ended, population growth in Malaya (renamed Malaysia after the country gained independence in 1957) was sustained by natural increase due to a major increase of the fertility rate from 1947 to 1955. Since at least the 1950s, life expectancy has rapidly increased and the mortality rate has significantly declined. This situation was due to the improvement of nutrition, the establishment of preventive health programs, and better accessibility to curative medicine. However, from the 1960s onward, the fertility rate has continued to decline to the present day. Since this chapter's projection follows assumptions based on historical fertility and mortality, it unsurprising that natural increase is expected to decline continuously in the future. Based on these trends, Malaysia will clearly be in the third phase of demographic transition during the projection period: both fertility and mortality decline continuously, and rural-urban migration maintain its dominance but less so than before.

Furthermore, this result confirms that future age structure (see Section 7.5.2 for more detail) will be influenced by the decline of natural increase. More adults or working-age people will survive, and this may significantly impact future economic growth resulting from an increasing number of labourers and fewer children to support, which relates to demographic dividend (Gribble & Bremner, 2012).

7.7 Settlement projections: Migration flows in 2015-2040

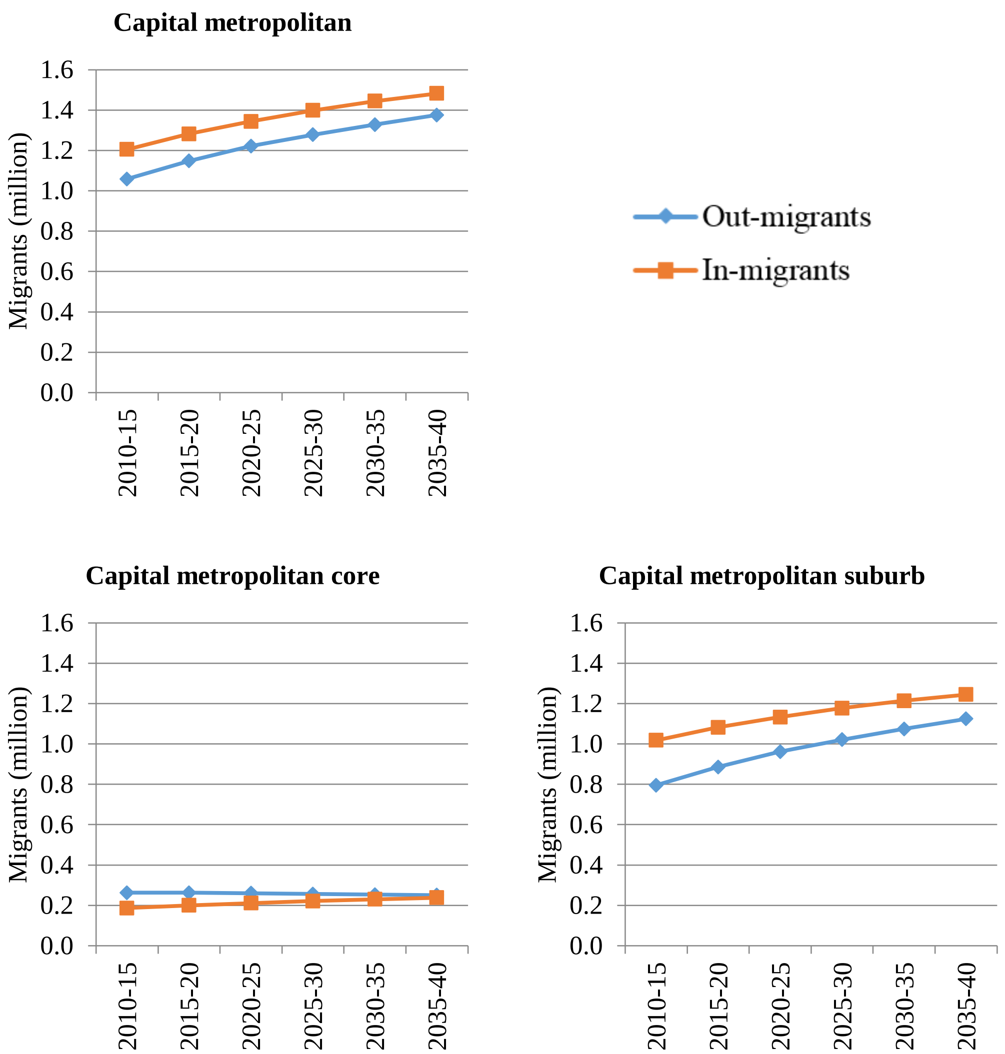
Many approaches to projecting future migration flows can be taken. The approach adopted here is to identify current district-level age-specific rates of in- and out-migration and to assume that these rates will remain constant over time. This approach projects that in- and out-migration flows will fluctuate in response to changes in the population size of origin and destination districts. The population of a district influences the size of its out-migrant flow, while the population of that district relative to all other districts influences it shares of the total pool of out-migrants received as in-migrants. Some of these calculations are conducted separately for each age group, as the overall inflows and outflows for a district are sensitive to changes in age structure.

The approach outlined was adopted due to (a) the desire to consider the influence of changing age structure on future migration flows and (b) data limitations. An alternative approach would be to extrapolate future rates based on historically observed trends for these rates, but this was rejected because simple extrapolation leads to unrealistically extreme future scenarios, while more sophisticated approaches allied to structure modelling have data requirements that could not be met. A second alternative approach also considered was the application of a spatial interaction model of the kind presented in Chapter 6. However, projecting a range of socio-economic variables, such as district-level unemployment rates, into the future is itself a hazardous exercise and would lie beyond the scope of this thesis. Further, the implementation of a simple interaction model based on origin and destination population sizes led to implausible flow distance and migration scenarios for the future. Full details of the approach that was actually adopted is outlined in the remainder of this section. It provides a projection of what will happen if current patterns of migration persist.

As mentioned in the methods section, this chapter projects only internal migration and excludes international migration (see Section 7.3.3 and Section 7.3.6 for a detailed explanation). This section is divided into four sub-sections: internal migration flows in Section 7.7.1, internal migration flows by age group in Section 7.7.2, net migration flows in Section 7.7.3, and net migration flows by age group in Section 7.7.4.

7.7.1 Internal migration flows

Internal migration is defined here as involving a change of place of residence that crosses a district boundary. For this study context, movement from one settlement type to another type is examined by comparing the number of in-migrants and out-migrants in each settlement types for the projection period of 2010-2040.



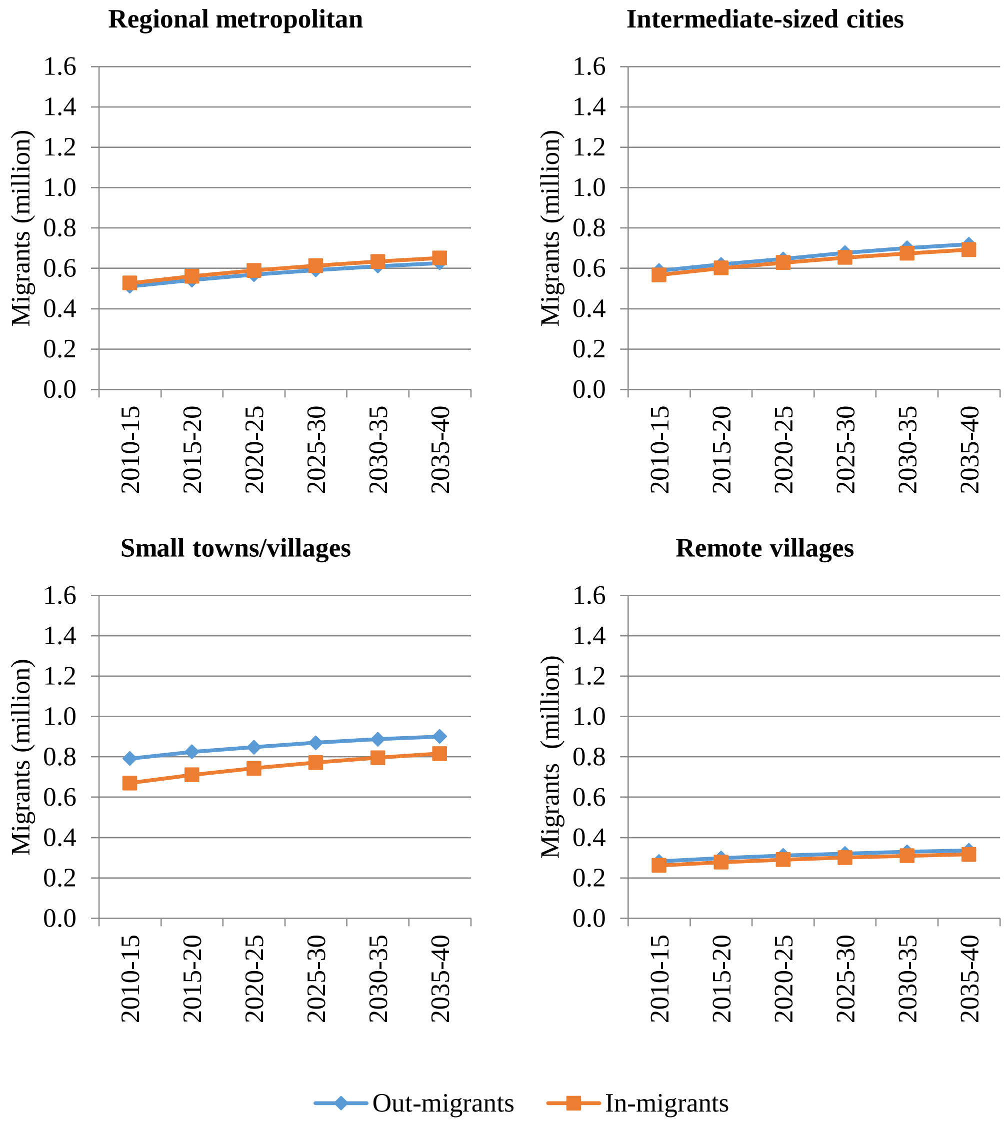


Figure 7. 17: Number of in-migrants and out-migrants by settlement type, 2010-2040

There are three findings based on Figure 7.17. First, the capital metropolitan area is projected to have net in-migration throughout the projection period. However, the capital metropolitan core and capital metropolitan suburban areas display a different migration pattern. The capital metropolitan core is projected to move from net out-migration flows to a balance of flows, while capital metropolitan suburban areas are projected to have net in-migration flows. Second, small towns/villages are projected to have net out-migration flows in the future. Finally, migration flows in the regional metropolitan areas, intermediate-sized cities, and remote villages are projected to be more or less balanced.

Because migration flows are projected from the sum of in-migrants and out-migrants instead of out-migration flows from origin to destination (see Section 7.3.6 for more detail), it is impossible to identify flows between settlement type (e.g., rural-urban, urban-rural, urban-urban, or rural-rural), only the number of in- and out-migrants in each district/settlement type. However, since the projections are based on recent migration trends (e.g., from 2005-2010), which are assumed to remain constant into the future, future migration patterns are foreseeable. The large concentration of migrants in capital metropolitan areas (especially in suburban areas) is likely due to urban-urban migration as well as rural-urban migration. In contrast, rural-urban migration continues to dominate the flows in all other settlement types. Future migration patterns are strongly related to the hypothesis in mobility transition theory (Zelinsky, 1971). The theory combines all components of change (births deaths, and migration) to explain demographic changes and the migration process within a framework. Malaysia may undergo Phase III of vital and mobility transition based on major rural-urban migration in most settlement types, a major decline of fertility rates (See Section 7.6.1), continuous decline of mortality rates (See Section 7.6.2), and significant decline of natural population increase (See Section 7.6.3).

Based on the existing population and migration studies (in Chapter 4 and Chapter 5, respectively) and projection results, rapid suburbanisation is expected to occur continuously if recent migration trends persist in the future. However, the long suburbanisation process in the capital metropolitan area (until 2040 at least) may worsen the current conditions of the city. Urban sprawl has been a major issue in Malaysia since the 1980s due to rapid suburbanisation in metropolitan regions caused by uncontrolled urban development growth (Abdullah, 2012; Hasan & Nair, 2014). According to Abdullah et al. (2009), urban sprawl occurs when urban development grows faster than population growth. Furthermore, a new inter-city rail project (The East Coast Rail Link, or ECRL) announced in 2016 that will connect cities (including cities in capital metropolitan suburban areas) is likely to further stimulate urban growth and development along the rail corridor (Malaysia Rail Link, 2019). The increase of large-scale urban development projects and the concentration of migrants in metropolitan cities, as well as the urban sprawl problem, may have a major impact on urban development growth and pressure the Malaysian government to provide more expenditure for housing, infrastructure, and amenities. Therefore, it would be best for the Malaysian government to consider controlling urban development growth in the capital metropolitan area as well as discouraging rural-urban migration through the implementation of additional policies.

7.7.2 Internal migration flows by age group

Age plays an important role in determining migration flows. Generally, young adults migrate more often to areas that offer more economic opportunities (e.g., high wages, employment opportunities). To examine migration patterns by age group, Figure 7.23 and Figure 7.24 compare the number of in-migrants and out-migrants by age group and settlement type for the first and final projection periods, 2010-2015 and 2035-2040.

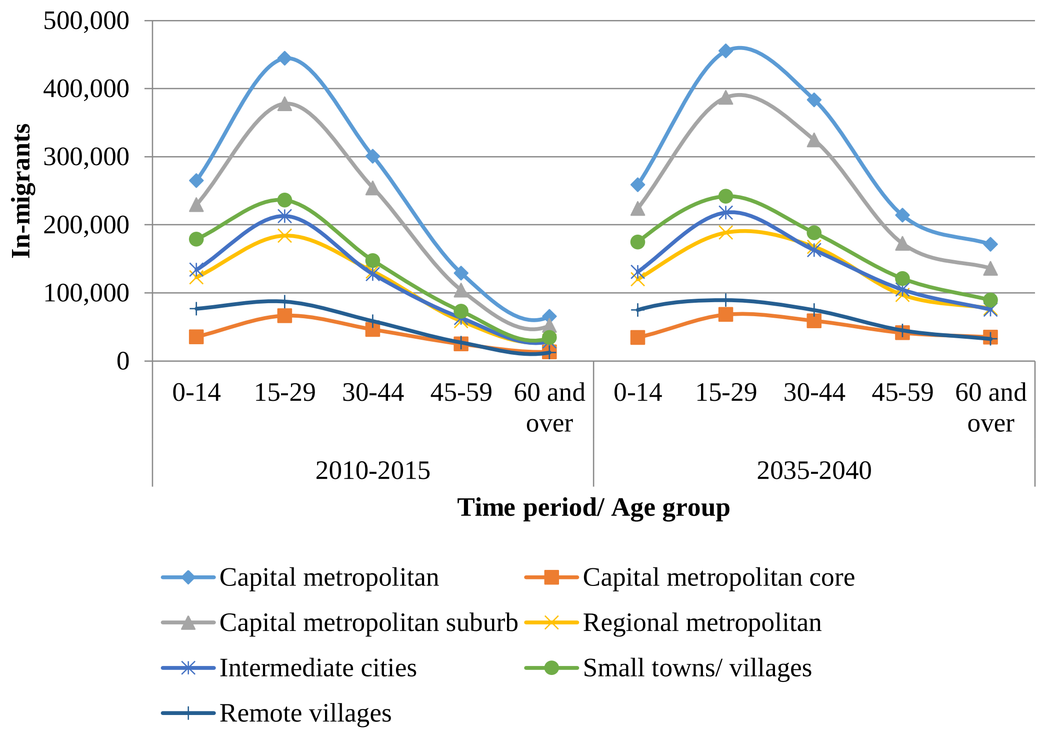


Figure 7. 18: Number of in-migrants by age group and settlement type, 2010-2015 and 2035-2040

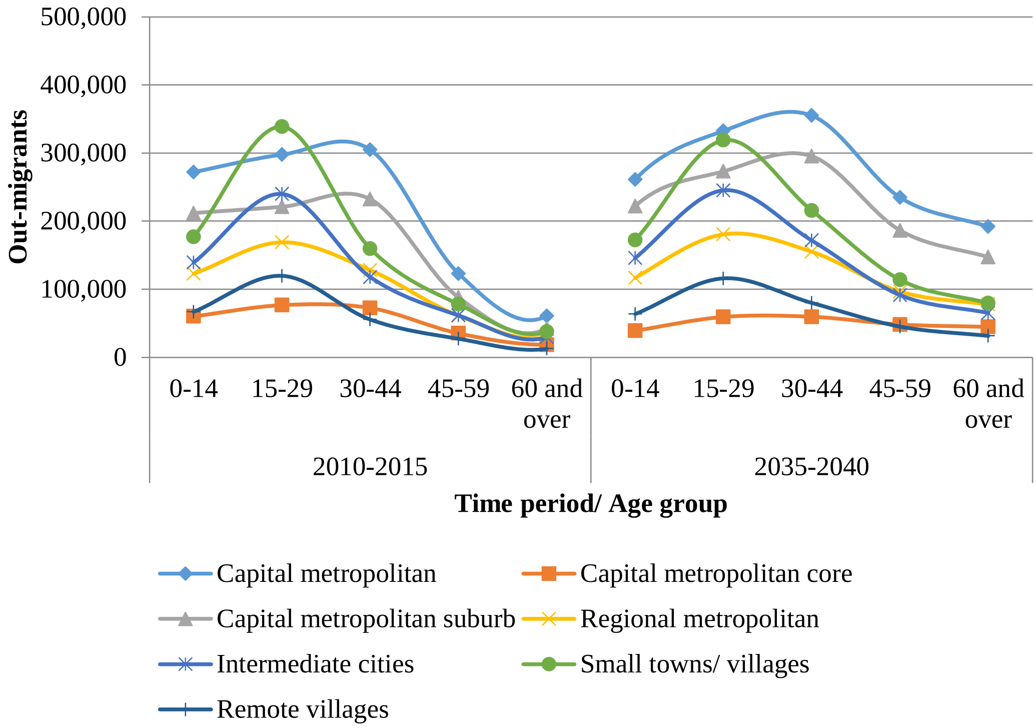


Figure 7. 19: Number of out-migrants by age group and settlement type, 2010-2015 and 2035-2040

As shown in Figures 7.18 and 7.19, there are similar patterns for both in-migrants and out-migrants in most settlement types for the first and final periods. Young adults (ages 15-29) account for the most migrants while the elderly account for the least in the initial period. However, in the final period, there is a significant increase in the number of older adult migrants (middle-aged adults, mature adults, and the elderly). In contrast, there is little change in the number of child/adolescent and young adult migrants. The major increase in the number of older adult migrants in the future relates to the decline in mortality and increase in life expectancy in the future (see Section 7.6 for more detail), which will to more migration. Common factors of migration include marriage commitments, changing jobs, and retirement (Bernard et al., 2014). The migration models in the previous chapter also capture similar factors, which supports that previous migration flows in Malaysia (e.g., 1980-2010) were influenced by marital status and changing jobs.

Despite the similarities, there are also differences in migration patterns by settlement type. First, while most age groups are projected to experience increased flows in the future, this will decline for some age groups. The number of out-migrants ages 29 and below in the capital metropolitan core will decrease because there will be fewer people of this age living in the city than in the initial period, hence fewer will out-migrate in the future. This is due to the continuous major decline of fertility rates and population ageing. Second, the number of middle-aged adult out-migrants will exceed the number of young adult out-migrants in capital metropolitan suburban areas. Similar to earlier explanations, the low fertility levels in larger settlements will lead to a smaller population ageing into young adults. On the other hand, more people will survive and become middle-aged adult, resulting in larger total flows of middle-aged adults in the future compared to young adults.

To compare in-migration and out-migration flows, the next section will discuss net migration flows.

7.7.3 Net migration flows

The difference between in-migration and out-migration is known as net migration. Figure 7.20 shows projected net migration for all settlement types in Malaysia, 2010-2040. Note that positive net migration indicates that there are more in-migrants than out-migrants, while negative net migration indicates the opposite.

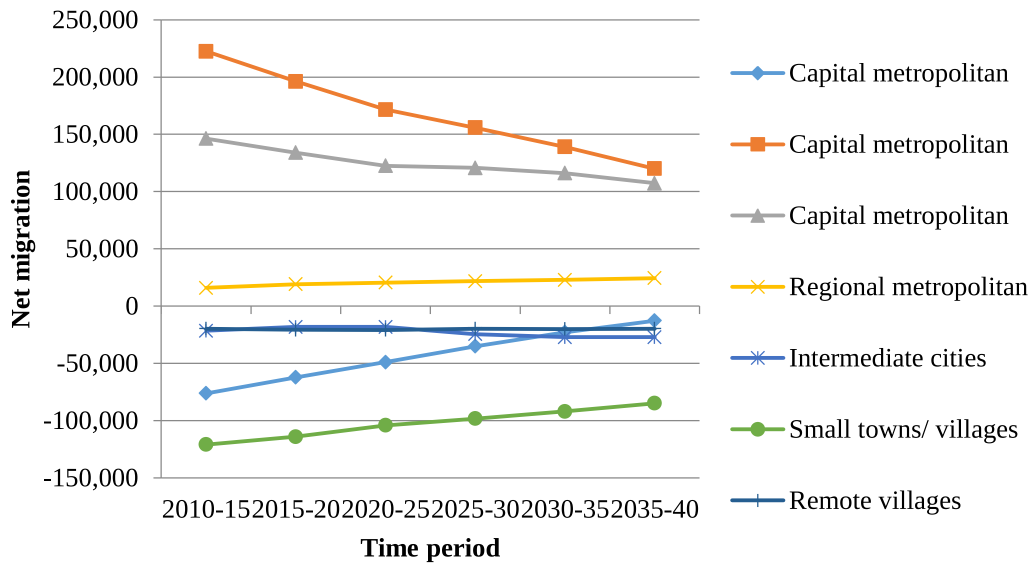


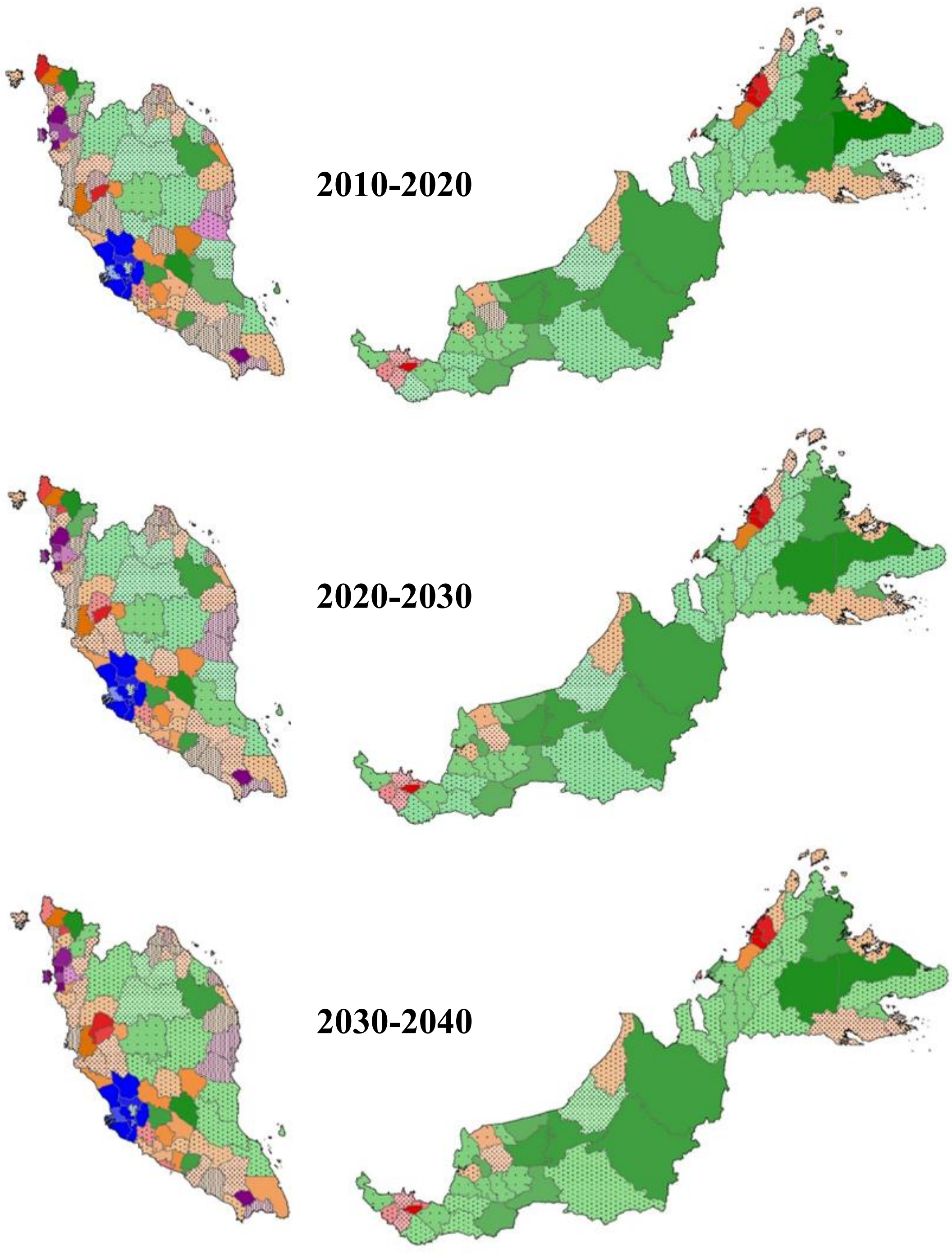
Figure 7. 20: Net migration by settlement type, 2010-2040

The results in Figure 7.20 show that metropolitan cities (capital metropolitan and regional metropolitan areas) are the only settlements projected to experience net-inflows. In other words, these cities will have more in-migration than out-migration flows in the future. However, within the capital metropolitan area, there will only be net overall in-migration because net migration into the suburbs will offset net-outflows from the core. For all other settlement types, the results shows a net outflow, possibly due to ongoing urban-urban and rural-urban migration. Small towns/villages are expected to have the largest net-outflows, albeit at declining levels over time. This is due to the diminishing influence of migration into larger cities, resulting in fewer people moving away from small towns/villages. The net-outflows for intermediate-sized cities and remote villages, on the other hand, will see no major changes in the future.

Related to differential urbanisation theory, these results suggest that Malaysia will still in the APC stage through 2040 from the dominance of net-inflows into capital metropolitan suburban areas. However, as time passes, the net-inflow of this settlement type will gradually decrease, signifying a slowdown in the suburbanisation process. This is possibly due to the decline in the number of people of active migrant age (particularly young adults) due to major declines in fertility. To validate this assumption, the next section examines net migration by age group.

This contrasts with the historic flows (1980-2010) shown in Chapter 5, where the net- inflows into capital metropolitan suburban areas gradually increased throughout the period. This is because different migration inputs were used for each chapter; data on Malaysian citizens was used for this chapter while data on the overall in population Malaysia was used to examine internal migration in Chapter 5. Nevertheless, Malaysian citizens contribute to the majority of internal migration flows since they constituted 92 percent of the overall population in 2010 and, according to official projections, will retain this dominance in 2040 with 91 percent. Hence, the projections of the net migration of Malaysian citizens in this chapter may be sufficient for explaining future net migration of the overall population of Malaysia.

Furthermore, another important finding is that the net-inflows in regional metropolitan areas will gradually increase in the future. This situation resembles the differential urbanisation theory hypothesis that states Malaysia is likely moving towards the polarisation reversal stage based on the shrinking dominance of capital metropolitan suburban areas and projected increases of net-inflows into regional metropolitan areas. However, the difference of net-inflows between these two settlements is quite large (more than 50,000 in 2040) and hence Malaysia is not expected to enter the polarisation reversal stage until shortly after 2040.



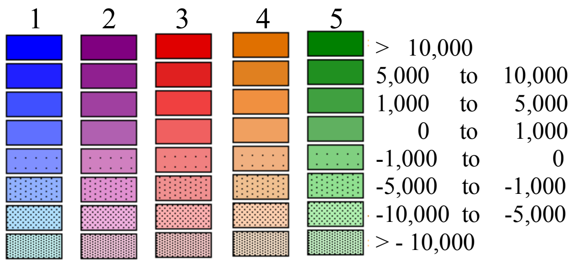


Figure 7. 21: Maps of net migration by settlement type, 2010-2040

1) Capital metropolitan area
2) Regional metropolitan areas
3) Intermediate-sized cities
4) Small towns/villages
5) Remote villages

7.7.4 Net migration flows by age group

This section expands upon the previous section by examining net migration by age group, comparing the first and final projection periods, 2010-2015 and 2035-2040 (Figure 7.22).

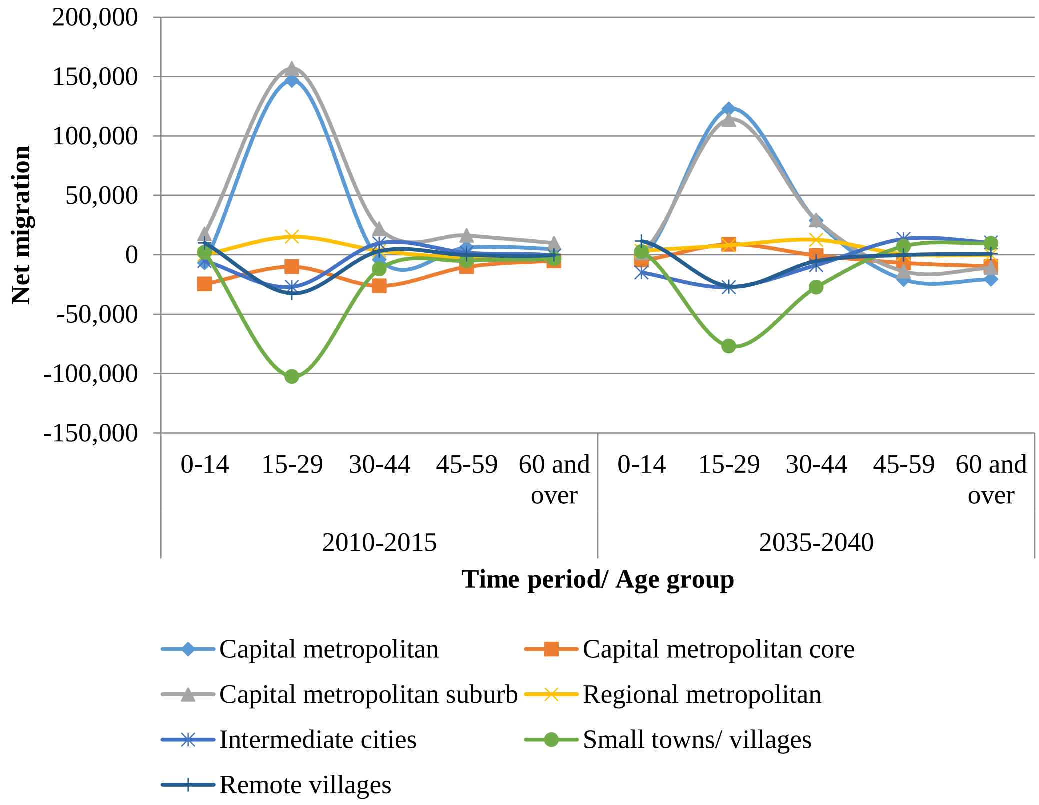


Figure 7. 22: Net migration by age group and settlement type, 2010-2015 and 2035-2040

Based on Figure 7.22, net migration in capital metropolitan suburban areas and small towns/villages will see major change in population and contradictory patterns (particularly for young adults). It is projected that more young adults will be moving into capital metropolitan suburban areas than moving out; in contrast, more young adults will be moving out of small towns/ villages than moving in. However, the size of net migration will decrease in the final period.

Although this analysis does not show the net flows between settlement types, the contradictory net migration patterns of capital metropolitan suburban areas and small towns/villages indicate a classic rural-urban migration pattern. Apart from rural-urban migration, the influx of young adults in capital metropolitan suburban areas is also due to urban-urban migration. These results are similar to those in Chapter 5; urban-urban migration was more dominant than rural-urban migration in the capital metropolitan in 2000 and will continue to be since migration projections are assumed to be constant in the future. Another surprising finding is the change from net-outflows to net-inflows for young adults in the capital metropolitan core. This result possibly relate to the increasing growth of fertility in this settlement type (see Figure 7.14) and an ageing population, hence increasing the number of young adult migrants in the future.

7.8 The relative importance of natural increase and net migration flows

Natural increase and migration flows, which are discussed in Section 7.6 and Section 7.7, respectively, are the main components of future population change. This section examines their relative contributions to future population change and the overall population. However, note that this thesis does not examines the contribution of future net international migration which also causes population change. Although future net international migration is excluded, their results would be similar to future net internal migration because 1) the size of international migrant stocks are relatively small from the overall population in 2010 (8 percent) (UNICEF, 2014), and 2) since internal migration is more dominant than international migration (UNESCO et al., 2012), therefore it should be sufficient to explain the overall story.

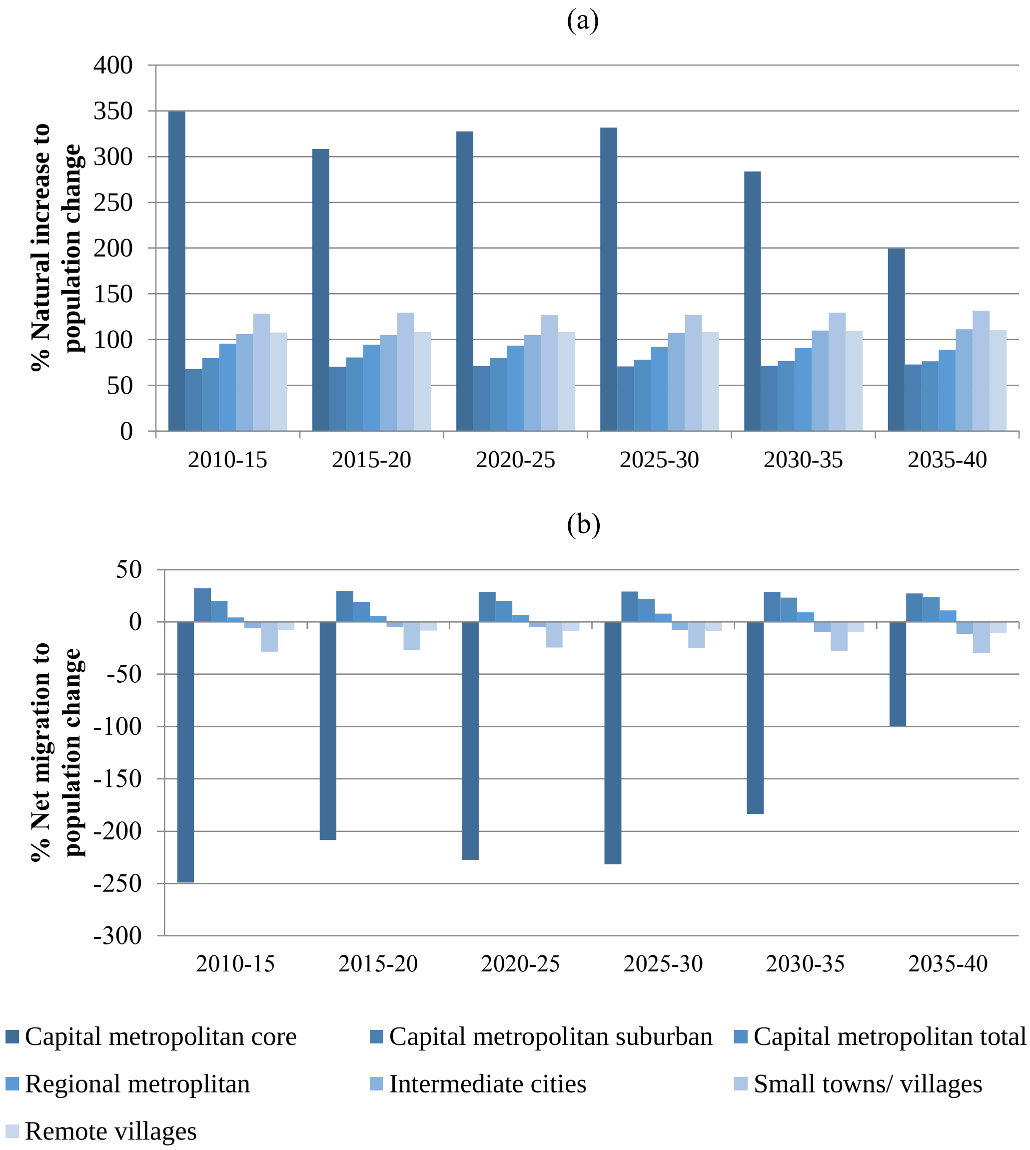


Figure 7. 23: The contribution of natural increase (a) and net migration (b) to population change

As shown in Figure 7.23, natural increase will have a larger contribution to population change than net migration for all settlement types in all projection periods from 2010-2040. Future population growth will be entirely due to natural increase in the capital metropolitan core, intermediate-sized cities, small towns/villages, and remote villages. For capital metropolitan suburban and regional metropolitan areas, natural increase will be the main driver of future change, but net migration inflows will also play a part.

Besides population change, another way to observe the influence of natural increase and migration flows is to examine their contribution to the overall population in Malaysia.

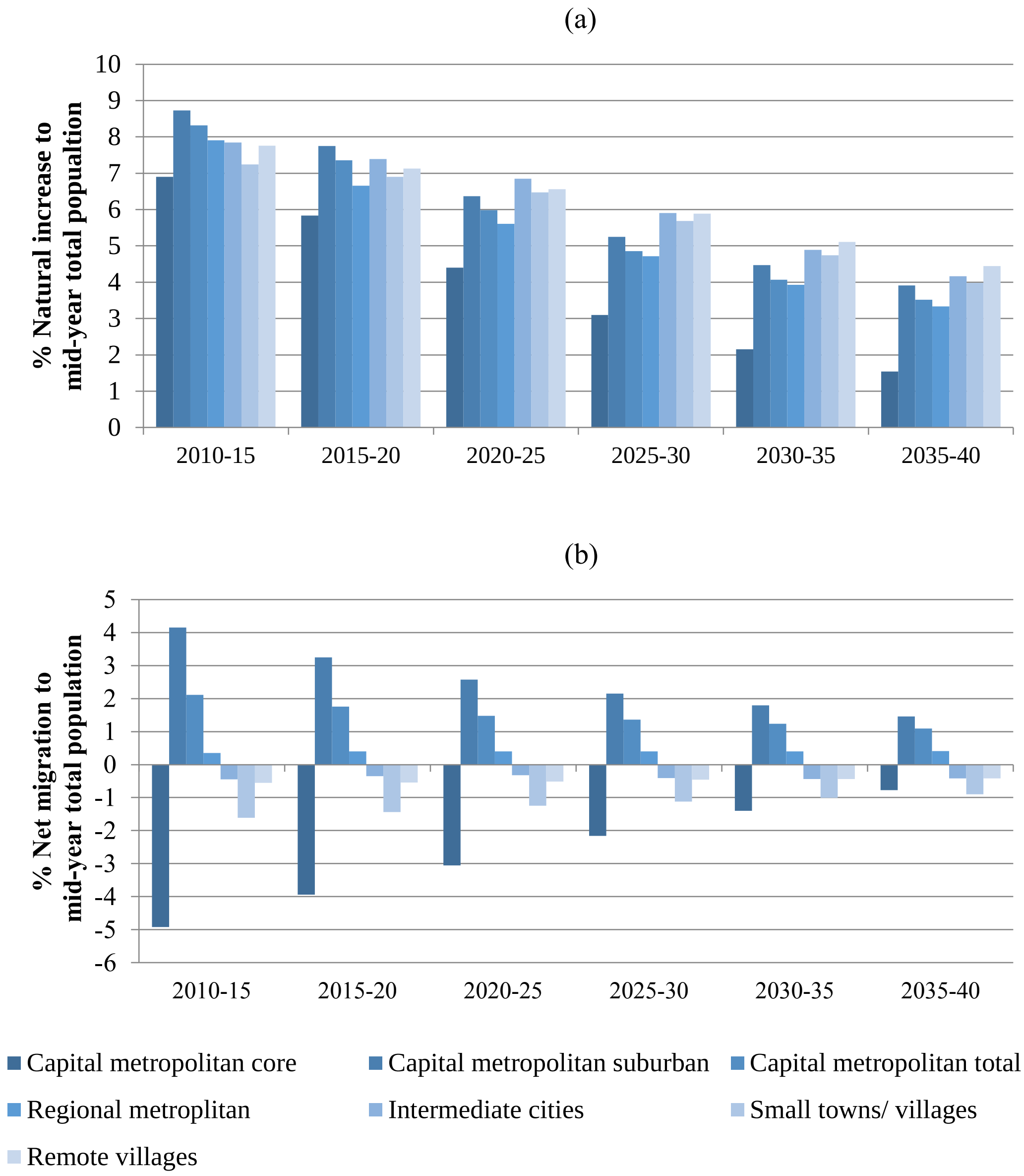
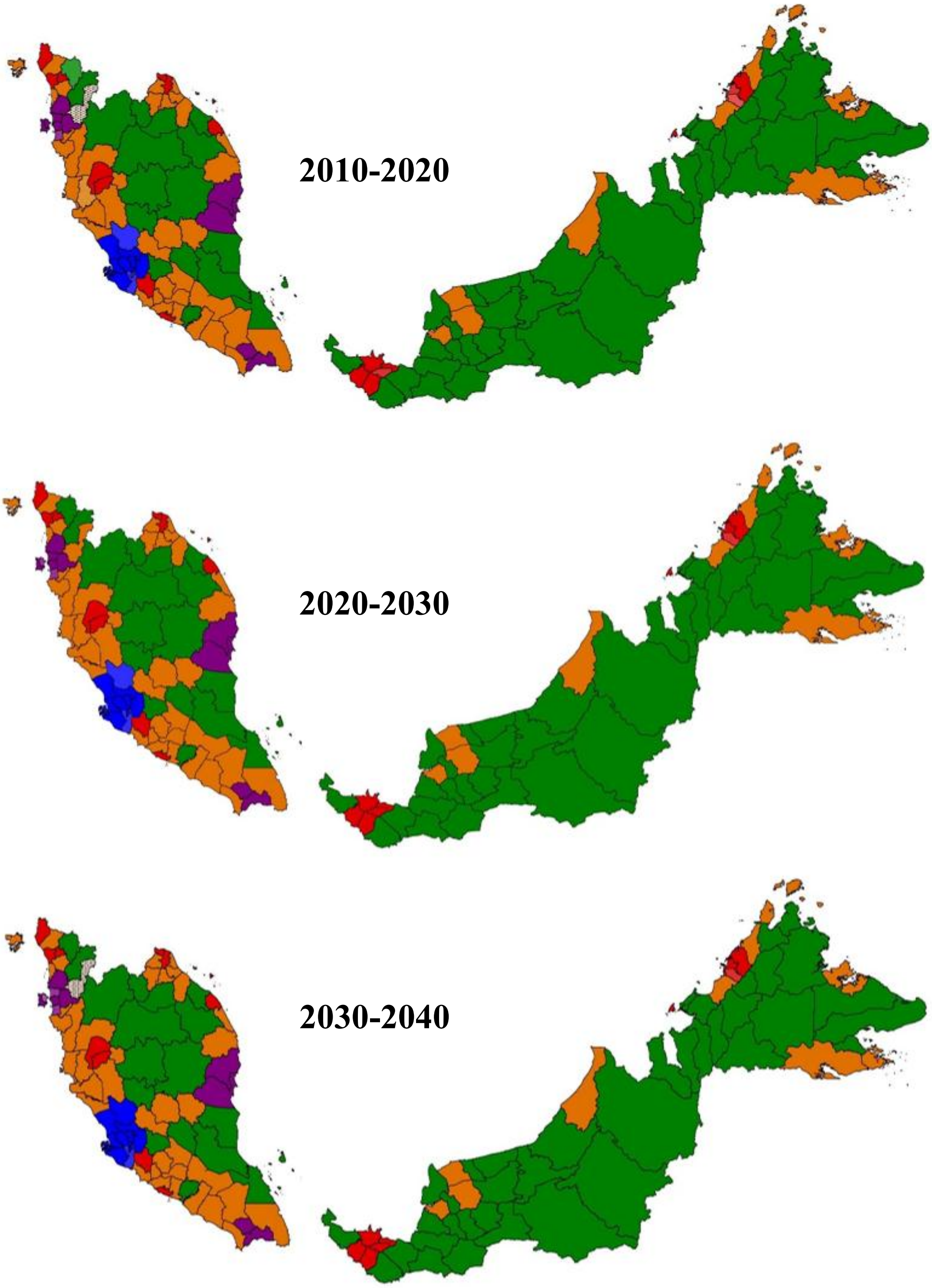


Figure 7. 24: The contribution of natural increase (a) and net-inflows (b) to total population (mid-projection period)

Figure 7.24 can be interpreted similarly as Figure 7.23 but more clearly highlights the scale of natural increase and net migration. Natural increase makes a larger contribution by percentage than net migration to the total projected population for all settlement types. However, this will gradually decline in size over time relative to the total population.

The major contribution of natural increase towards future population growth in all settlements relates to the age structure of the population. The large proportion of the female Malaysian citizens of reproductive age will result in more births despite a fertility decline to below replacement level: from 2.6 in 2010 to 1.9 in 2040. This phenomenon is known as population momentum. Population momentum arises when previous high fertility results in a large female population of reproductive age, hence leading to a large number of births (Keyfitz, 1971). According to Blue and Espenshade (2011), for countries still in the process of demographic transition and with a large proportion of females of childbearing age, or children who will enter reproductive ages, population momentum will have a significant impact on future population growth. On the contrary, for countries that have completed demographic transition, with a large proportion of the population ageing or elderly, alongside low fertility, population momentum will not have a significant impact on future growth and the total population will eventually decrease (Andreev; Kantorová; Bongaarts, 2013).

In the Malaysian context, according to Faizah (2007), from the late 1950s to 2006, the country was in the second stage of demographic transition due to a rapid decline of mortality and moderate decline of fertility. However, Malaysia will enter the third stage of demographic transition due to the continuous decline of mortality and fertility levels within the projection periods. There are two possibilities if fertility, mortality, and migration trends persist through 2040: 1) population momentum will still significantly contribute to population growth because of the high proportion of children who will be of reproductive age and females already of reproductive age (see Figure 7.2); and 2) the country may enter the fourth stage of demographic transition in which the fertility rate becomes the same as the mortality rate (see Figure 7.25 (a), which demonstrates the declining influence of natural increase on population growth).



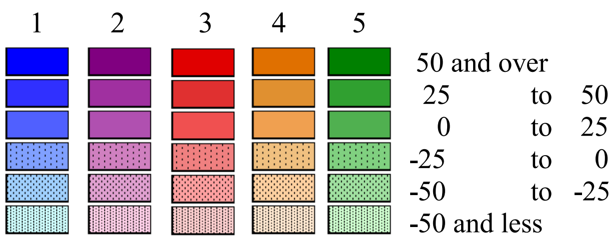
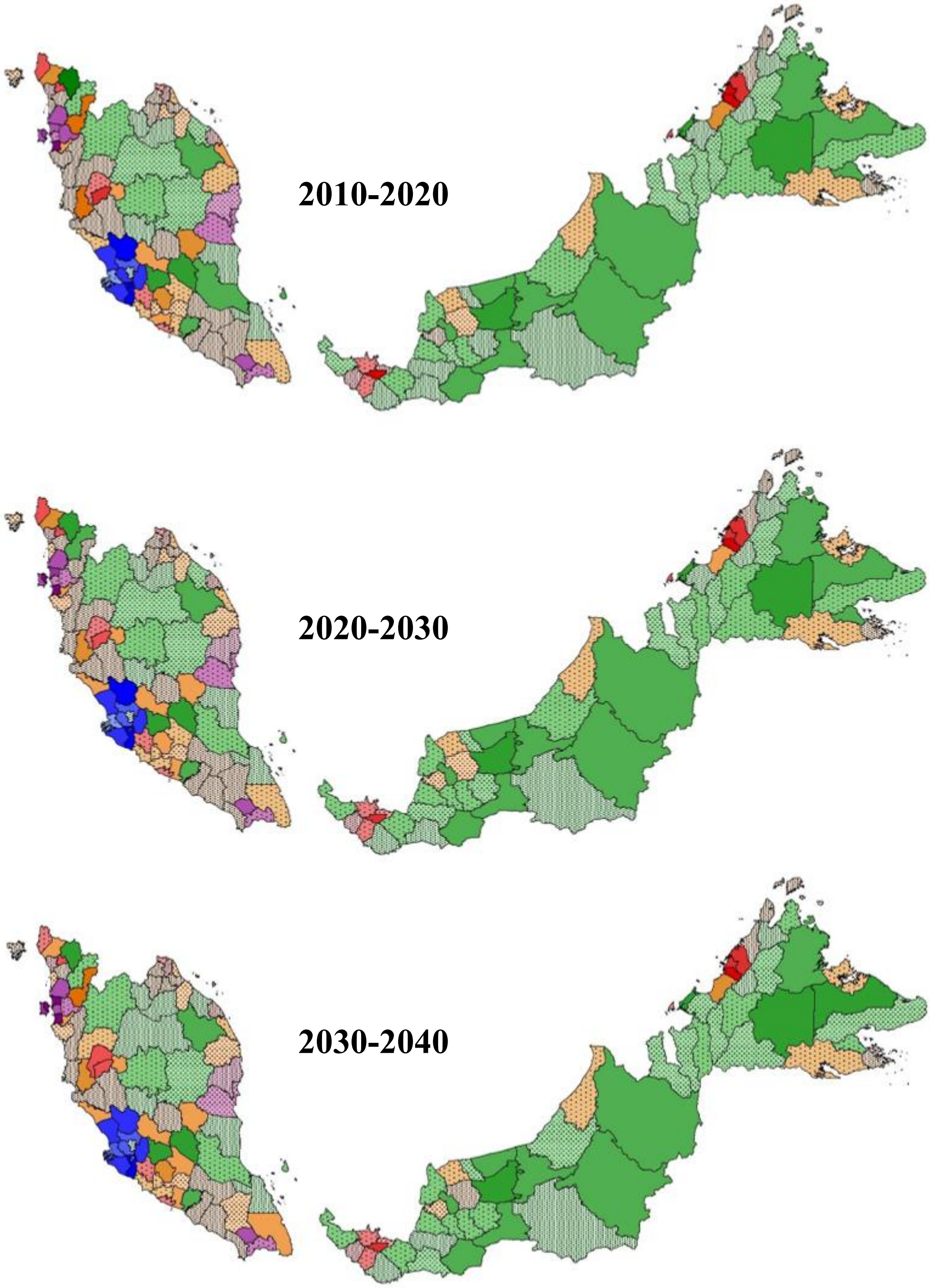


Figure 7. 25: The contribution of natural increase to population change, 2010-2040 (%)

Note:
1) Capital metropolitan area
2) Regional metropolitan areas
3) Intermediate-sized cities
4) Small towns/villages
5) Remote villages



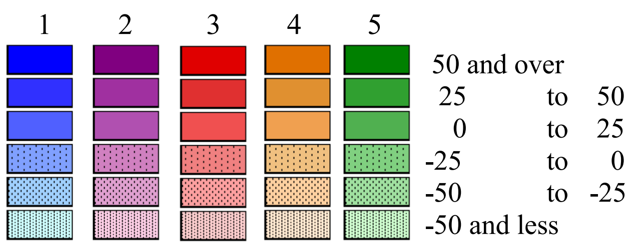


Figure 7. 26: The contribution of net migration to population change, 2010-2040 (%)

Note:
1) Capital metropolitan area
2) Regional metropolitan areas
3) Intermediate-sized cities
4) Small towns/villages
5) Remote villages

7.9 Conclusion

There are six important findings in this chapter. First, this chapter's projections match the official projections at the national level and are fairly similar to those at the state level. The difference between state projections is unsurprising mainly due to the different data and methods used to project internal migration. Despite the differences, both projections follow a similar concept: both projections assume current migration rates will unchanged in the future. However, this chapter's projections involve small-area unit flows (districts) rather than the large-area units (states) used for the official projections. This chapter's projection allows a detailed assessment of internal migration flows by not only district level but also settlement type, which is essential for comparing projected future urbanisation patterns to the urbanisation sequence hypothesized in differential urbanisation theory (from urbanisation to polarisation reversal to counterurbanisation).

Second, the population in all settlement types is projected to grow in size. However, capital metropolitan suburban areas are projected to have the largest population change in the future. In contrast, the future population in the capital metropolitan core is projected to diminish continuously. Apart from the capital metropolitan core, the shares and growth in small towns/villages are also projected to decline continuously in the future. Regional metropolitan areas and intermediate-sized cities, on the other hand, will see no major changes in relative standing, remaining stable throughout the projection period. The change of population in these settlement types will mainly be caused by natural increase rather than internal migration. There are fewer births in large settlements (especially the core city) than in smaller settlements and rural areas while deaths are more prevalent in large settlements than smaller settlements. The contrasting birth pattern between settlement types indicates that fertility levels are lower in large settlements than in lower-level settlements and rural areas. Generally, this is because of increasing accessibility and improvement of the social and economic conditions of the female population (e.g., availability of modern contraceptives, postponement of childbearing and marriage, increasing abortion rate, higher levels of education and employment among women) (Ernestina, 2002; Hirschman, 1980). Other reasons include the tendency to have more children in smaller settlements than in larger settlements due to location and housing type preference (Kulu, 2005; Kulu et al., 2009; Kulu & Vikat, 2007). In terms of mortality, more deaths are predicted in large settlements than in smaller settlements and rural areas because of the increasing prevalence of heart-related diseases and traffic accidents (Malaysia Department of Statistics, 2018). Living in big cities, especially in capital metropolitan areas, is quite stressful due to factors such as high living costs, high traffic volume, and high crime (Free Malaysia Today, 2017). Furthermore, the capital metropolitan area is known to have the high fatalities fur to traffic accidents since it is the busiest city in Malaysia. It is unsurprising that the city has a large number of deaths (Kunasekaran, 2017).

Third, all settlements will experience a major change in age structure in the future. The change in age structure between the starting and final year projections indicate that future population growth will mainly be sustained by adults (particularly adults ages 30 and over) or the working-age population (ages 15 to 65). This is due to a large proportion of the female population being of reproductive age, resulting in more births, despite fertility declines to below the replacement level: from 2.6 in 2010 to 1.9 in 2040. This phenomenon is known as population or demographic momentum. Population momentum arises when previous high fertility results in a large increase in the size of the female population of reproductive age, hence leading to more births (Keyfitz, 1971). The increase in the working-age population may have a significant impact on economic growth in the future, which relates to demographic dividend. The demographic dividend occurs when economic growth is accelerated by the decline of fertility and mortality, leading to a change in the age structure (Gribble & Bremner, 2012). As time passes, the proportion of children grows smaller than the working-age population. Hence, there are fewer children to support, leading to significant economic opportunities from the implementation of policies and investment (Gribble & Bremner, 2012).

Fourth, the projections capture a similar trend to that postulated in demographic transition theory and mobility transition theory. Malaysia may undergo Phase III of demographic and mobility transition, which involves the continuous decline of fertility and mortality, and major rural-urban migration in most settlements (except the capital metropolitan area). There are two possibilities if fertility, mortality, and migration trends persist through 2040: 1) population momentum will still significantly contribute to population growth because the proportion of children who will be at reproductive age and females already at reproductive age in 2040 will remain large (see Figure 7.2); and 2) the country may enter the fourth stage of demographic transition and mobility transition in which the fertility rate becomes the same as the mortality rate, coupled with a major increase in urban-urban migration.

Fifth, rapid suburbanisation will result if recent migration trends persist in the future. However, the long suburbanisation process in the capital metropolitan area (until 2040 at least) may worsen current conditions in the city. Urban sprawl has been a major issue in Malaysia since the 1980s due to rapid suburbanisation in the metropolitan regions caused by uncontrolled urban development growth (Abdullah, 2012; Hasan & Nair, 2014). The increasing number of large-scale projects and concentration of migrants in metropolitan cities, as well as urban sprawl, may have a major impact on urban development growth and pressure the Malaysian government to provide more expenditure for housing, infrastructure, and amenities. Therefore, it would be best for the Malaysian government to seriously consider controlling urban development growth in the capital metropolitan area as well as discouraging rural-urban migration through the implementation of additional policies.

Finally, in terms of differential urbanisation theory, Malaysia will remain in the APC stage until 2040 due to the dominance of positive net-inflows in capital metropolitan suburban areas. However, as time passes, the net-inflows into this settlement type will decrease, indicating a slowdown in the suburbanisation process. Net-inflows in regional metropolitan areas, on the other hand, will slightly increase over the years, indicating that this settlement type is slowly receiving a higher concentration of migrants. This situation resembles the differential urbanisation theory hypothesis, whereby Malaysia is moving towards polarisation reversal stage based on the shrinking dominance of capital metropolitan suburban areas and promising increase in regional metropolitan areas. However, the difference of net-inflows between these settlements is projected to remain relatively large, hence polarisation reversal is not expected to happen until shortly after 2040.

Chapter 8

Discussion and final conclusion

8.1 Introduction

The aim of this thesis is to investigate recent and future population growth, internal migration, and urbanisation in Malaysia from 1980 until 2040. In Chapter 1, five objectives were formulated to help achieve this goal. These objectives were then addressed in the chapters that followed through the application of relevant literature and methods. This chapter concludes all of the work done in this thesis by summarising the research findings in Section 8.2, presenting a discussion and limitations of the research in Section 8.3, and making recommendations for future work in Section 8.3.4. Overall, this thesis has successfully addressed its stated aim and objectives. More importantly, the thesis makes several novel contributions:

i. Rare application of the differential urbanisation theory in a developing country context and for the first time in Malaysia.
ii. Creation of a new settlement type in Malaysia that can be applied consistently from 1980 onwards and is compatible with the differential urbanisation theory.
iii. A detailed analysis of socio-economic drivers of internal migration in Malaysia by the application of smaller geographical units and the consideration of numerous socio-economic factors and types of flows.
iv. The first settlement type and district-level projections of Malaysia's future population using assumptions equivalent to those included in the official national and state-level projections.

8.2 Summary of findings

The summary of research findings of this thesis is based on the five research objectives formulated to achieve the over-arching aim of investigating recent and future trends of population growth, internal migration, and urbanisation in Malaysia from 1980 to 2040.

8.2.1 To review existing theoretical perspectives on population growth, internal migration, and urbanisation.

To understand urbanisation in Malaysia, it is necessary to first review what is already known about this process as it operates elsewhere. Therefore, Chapter 2 reviews what is currently known about the nature, causes, and consequences of urbanisation and demographic change, whether in the developed or developing world, to identify which aspects can be applied to the Malaysian experience. The conclusion reached is that differential urbanisation theory provides a useful lens through which to view all of these processes. The theory draws together literature and experience on urbanisation, polarisation reversal, and counterurbanisation into one over-arching theory. The transition between urbanisation stages can be identified by a 'clean break', when relative sizes of net migration flows between settlement types change, resulting in a change in urbanisation patterns (Champion, 2005). For example, urbanisation is evident when the net migration of large cities exceeds that of medium and small cities; polarisation reversal happens when net migration of medium and small cities exceeds that of the largest cities; and counterurbanisation happens when net migration of small cities exceeds that of large and medium cities.

To date, the theory has been applied for many developed countries but few developing countries (India, South Africa, Botswana, and Turkey). While most developed countries have experienced counterurbanisation, only two developing countries have reached that stage. Overall results for Malaysia in this thesis follow the same urbanisation pattern but not polarisation reversal and counterurbanisation from 2010-2040. This is to be expected because the largest city in Malaysia, the capital metropolitan area, maintains its dominance, especially in their suburbs, although this is predicted to diminish over time in the future. Although the country has the potential to move towards the next deconcentration stage, polarisation reversal, if rapid suburbanisation persists until at least 2040, this may distort the process. Similar to India, there is no clear evidence that the country will experience counterurbanisation due to the inadequacy of physical infrastructure, and the lack of institutional capacity may result in re-urbanisation instead of counterurbanisation (Jain et al., 2013).

8.2.2 To develop a new urban-rural classification of Malaysia based on urbanisation theory.

To test differential urbanisation theory, Chapter 3 develops a new urban-rural classification for Malaysia based on a theoretical approach and assumptions. This new classification is required because existing urban-rural classifications of Malaysia are not suitable. Although carefully constructed by various government agencies, existing classifications are different in terms of the number of urban and rural areas, physical boundaries, definitions, and measurements used, and they do not naturally tie in to differential urbanisation theory. More importantly, only basic information (e.g., total population and total migration) is available for the existing urban-rural units. To overcome this obstacle, this research uses alternative small-area units (districts and *mukim*), which are then transformed into urban-rural units. The comprehensive demographic information needed to test the theory in detail is available for these units for successive censuses on spatially consistent bases. However, there are no specific guidelines on how to differentiate the settlement type (i.e., large, medium, and small cities) except that they must be located independently from each other (Geyer & Kontuly, 1993).

To fit the theory approach in the Malaysian context, the first step is to classify the existing settlement hierarchy into new settlement types based on the theory approach (large, medium, and small cities). The national conurbation is categorised as the primate/largest city since this is known to be the largest settlement in the hierarchy; regional, sub-regional, state, and district conurbations are classified as intermediate-sized cities because they fit within regional, sub-regional, state, and district contexts; major and minor settlement centres are classified as small cities because these settlements act as local towns and nearby villages; and remote villages are villages located far from the cities. Finally, districts/*mukim* are then classified to reflect the settlement types contained within their boundaries. For example, if a district/*mukim* contains part of the national conurbation (whether the existing boundaries and/or urban built-up areas), it is classified as the capital metropolitan area. For another example, if a district/*mukim* contains many major and minor settlement centres, it is classified as a small town/village. The new urban-rural classification and spatial units are then used to examine urbanisation in Malaysia in the subsequent chapters.

8.2.3 To investigate patterns of population growth, internal migration, and urbanisation in Malaysia over the recent period (1980-2010).

From a demographic perspective, urbanisation (the increase of the share of the urban population) is generally caused by natural increase of the urban population and rural-urban migration. However, due to limitations in terms of the availability of fertility and mortality information, an examination on the former cause is not included in this thesis. In contrast, migration information is available for small-area units, which are examined in Chapter 5. First, however, Chapter 4 provides an overview of overall population growth and urbanisation trends in Malaysia.

Over the period of 1980-2010, Malaysia as a whole grew rapidly but at a decreasing rate. All settlement types grew in size over this period. However, relative population shares changed over time. The capital metropolitan area had the highest population growth followed by settlements lower in the hierarchy. In other words, growth is in hierarchical order, where the larger the settlement, the larger the growth. However, the results on population shares tell a different story. The population share in larger settlements (capital and regional metropolitan areas) increased but decreased in smaller settlements. Note that these metropolitan cities are a combination of the core city and surrounding suburban areas. As expected, population growth in suburban areas in both the capital and regional metropolitan areas was significantly higher than in the core cities. In terms of population share, the core and suburban areas in both types of metropolitan areas display a contrasting pattern; the share in suburban areas significantly increased, offsetting a decline in the core.

In relation to differential urbanisation theory, Malaysia may have been in the final stage of urbanisation, APC, since 1980 as a result of rapid population growth in capital metropolitan suburban areas offsetting a decline in the capital metropolitan core (*Kuala Lumpur*). As the core city is saturated with economic and physical development, this causes agglomeration diseconomies and decentralisation towards peripheral suburban areas. Further, *Kuala Lumpur*'s confinement by limited physical space has led to urban expansion beyond its borders towards peripheral suburban areas. The analysis revealed a classical pattern, with the population in large metropolitan cities dominated by young adults (ages 15 to 29). There is no clear sign that Malaysia will shift towards the next deconcentration stage, polarisation reversal, because the capital metropolitan area (mainly suburban areas) will maintain dominance in terms of relative population size compared to other cities (intermediate-sized and small cities). However, the real test of the theory is the observation of migration between settlement types to identify migration patterns and their contribution to the urbanisation process

The analysis of internal migration in Chapter 5, however, denies the claim made in Chapter 4 based on total population and change alone. The results show Malaysia has not been in the APC stage since 1980 but rather in the former stage, IPC, based on high rural-urban migration in the largest city, the capital metropolitan area. However, during 1995-2000, the migration pattern changed from rural-urban to urban-urban migration, thus marking a transition in the urbanisation process from the IPC stage to the APC stage. Suburbanisation has taken place, replacing urbanisation due to large out-migration flows from the capital metropolitan core into capital metropolitan suburban areas. This situation means that the capital metropolitan core was no longer the main destination of migrants by 2000. The large net-outflows in the capital metropolitan core indicate that urbanisation was entirely due to natural increase instead of net migration. One of the main reasons for this is the urban sprawl phenomenon. Urban sprawl is evident from the establishment of many new townships and major highways situated outside *Kuala Lumpur,* which attracted more migrants.

In terms of age groups, migrants aged 45 and over had higher mobility than those of a younger age. These results are surprising because younger migrants (especially young adults) generally have the highest mobility, which then slowly declines with increasing age and sometimes increases again for parents with young children and those of retirement age (Bernard et al., 2014). A possible reason for this is the decline in the number of people of active migrant age (young adults) due to the continuous decline in fertility; fertility in Malaysia has been declining since the 1960s, and by 2010, the total fertility rate was at replacement level (2.1 births per woman). However, despite the high mobility of migrants aged 45 and over in most settlement types, the change in the number of migrants was relatively small. In contrast, the low mobility of migrants aged below 45 resulted in a significant decline in the number of migrants. Finally, all settlement types display a common age distribution, with young adults (age 15-29) comprising the largest groups of in-migrants and out-migrants.

If recent population and migration trends persist in the future, capital metropolitan suburban areas may continuously grow rapidly, while the primacy of the capital metropolitan core will be further eroded. Further, there is no clear sign that the country will move towards the next urbanisation stage, polarisation reversal, based on the dominance of the capital metropolitan area compared to smaller settlement types.

8.2.4 To identify and explain the determinants of Malaysia internal migration, 1980-2010.

As well as understanding the nature of recent migration patterns in Malaysia, an objective of this thesis is to offer an explanation of the observed migration flows. Chapter 6 is an extension of Chapter 5 and attempts to identify and explain the determinants of internal migration in Malaysia. Spatial interaction models were chosen as the main method for investigating these issues. There are two main reasons for this: 1) the models are able to explain and identify the best set of factors that influence the migration flow; and 2) the available migration data are for aggregate flows (e.g., net, in-, and out-migration of a district; and total flow from origin to destination by district), which rules out other approaches that require the use of individual-level data. To test the model comprehensively, two modelling approaches are proposed: 1) modelling total migration flows; and 2) modelling origin-destination flows. The total migration flow models captured only a few determinants (e.g., the net migration model captures one determinant) because this approach models only aggregate flows to each place, rather than the full set of individual origin-destination flows. This highlights that there is no such thing as net migration; internal migration consists of both in-migrants and out-migrants. Origin-destination flow models are able to explain migration more clearly than total migration flow models because they take account of inflows and outflows separately.

In summary, there is a clear distance effect, mediated by both crossing the sea between East and West Malaysia and flows between settlement types. The key drivers of flows (push factors) from origins are more adults (either young, middle-aged, or mature adults), single individuals, married individuals, ethnic majority, high-paid jobs, middle-paid jobs, secondary industry jobs, and tertiary industry jobs. Origins with more married people have more outflow, possibly because of the factor of following a partner/spouse. However, origins with more single people also have more outflow. This relates to the fact that there are more flows from origins with more young adults that are leaving school, completing higher education, or entering the labour force. Additionally, origins with more ethnic majority members have more outflows, which is related to the high population growth of the ethnic majority in all cities (see Chapter 4 in Section 4.5.2) and encouragement from policy implementation by the Malaysian government. Finally, the increase of flows at origins with many jobs is due to the migration of people seeking a similar or different type of job (see the discussion on jobs that pull more flow in the next paragraph).

In contrast, the key drivers of flows (pull factors) from destinations are more adults, the elderly, working adults, married individuals, ethnic majority, educated population (secondary and tertiary education), high-paid jobs, medium-paid jobs, secondary industry jobs, and tertiary industry jobs. These results are similar to those in Chapter 5, with older adults and elderly migrants having greater mobility than younger migrants. This is surprising because young adults typically have the greatest mobility, which then declines with increasing age (Bernard et al., 2014). A possible reason for this is the decline in the number of people of active migrant age (young adults) due to the continuous decline in fertility since the 1960s. Further, as explained, following family and marital status are known to be primary reasons for internal migration in Malaysia (UNESCO et al., 2012). The increase of flows to destinations with several types of jobs agrees with classical migration theory, which states there will be more migration into areas that offer higher wages or better jobs (Harris & Todaro, 1970). Further, the increased attractiveness of destinations that offer more medium-paid and secondary industry jobs reflects the economic transition from primary to multi-sector sector industries in the early 1980s. The manufacturing sector and modern services grew substantially and became centralized in the vicinity of cities (Abdullah, 2003).

8.2.5 To project future population growth, internal migration, and urbanisation in the period from young adults 2015 to 2040.

The analysis of three decades of historic population change (1980-2010) shows clear evidence of urbanisation but no evidence of polarisation reversal or counterurbanisation as postulated in differential urbanisation theory. Therefore, Chapter 7 examines the likely patterns of future population growth, internal migration, and urbanisation in Malaysia from 2015 to 2040, adopting official projections of trends in fertility, mortality, and migration disaggregated to district level. There were nine important findings in this chapter.

First, this chapter's projections match the official projections at the national level and are reasonably similar for the state level. The differences that arise for state projections are unsurprising, mainly due to the different data and methods used for projecting internal migration. Despite the differences, both projections follow a similar concept; both assume current migration rates will remain unchanged in the future. However, this chapter's projections involve flows of small-area units (districts) rather than the large-area units (states) used in the official projections. This chapter's projections allow a detailed assessment of internal migration flows by not only district but also settlement type, which is essential for comparing projected future urbanisation patterns to the urbanisation sequence hypothesized in differential urbanisation theory, which the official projections do not cover.

Second, in terms of fertility, the birth share is the largest in small towns/villages because this settlement type has the highest number of births. In contrast, the growth in births shows that most settlement types will experience a decline in negative growth until 2025, which will then diminish in subsequent years. In terms of change relative to the initial year, the capital metropolitan core is projected to experience the fastest and largest decline in the birth rate, similar to regional metropolitan and capital metropolitan suburban areas, albeit at a slower rate. For other settlement types, birth s will increase at a similar rate until 2025 but then decline in subsequent years. In other words, larger settlements are projected to see fewer births (especially in the capital metropolitan core) in the future while smaller settlements will follow the same trends after 2025. The contrasting patterns between larger and smaller settlements (including rural areas) indicate that those living in larger settlements have lower fertility than those living in smaller settlements. Generally, this is because of increasing accessibility to education and improvement of social and economic conditions of the female population (e.g., usage of modern contraceptive, postponement of childbearing and marriage, increasing abortion rate, higher level of education and employment among women) (Ernestina, 2002; Hirschman, 1980). The results also reflect the assumption made for the fertility rate used in the projection; fertility rates are assumed to decline continuously through 2040. The decrease of the fertility rate in Malaysia began in the 1960s and continues to decline to the present day. Furthermore, although fertility rates are projected to fall below the replacement level by 2040, the total population continues to increase in all settlement types, hence indicating a population momentum effect. Population momentum is a situation when previous high fertility results in a large proportion of the female population being of reproductive age, hence leading to more births (Keyfitz, 1971).

Third, in terms of mortality, the share of people dying in larger settlements will slowly increase but will decrease in smaller settlements. In other words, more people are projected to die in larger towns than in smaller towns. The growth and relative change results can be interpreted similarly. These results reflect the assumptions made about future life expectancy in the projections; it is expected to increase continuously in the future. According to Malaysia Department of Statistics (2018), the main cause of death in urban areas in Malaysia is ischaemic heart disease, which mostly affects people aged 41 and over; in contrast, traffic accidents are the main cause of death for those aged 40 and below. Living in big cities, especially in the capital metropolitan area (Kuala Lumpur and its suburban areas) is quite stressful (e.g., high living costs, high traffic, high crime) (Free Malaysia Today, 2017) and is less healthy, which can lead to heart-related diseases (e.g., respiratory disease and lung cancer) (O'Reilly et al., 2007). Furthermore, the capital metropolitan area has the highest number of fatalities due to traffic accidents since it is the busiest city in Malaysia. Therefore, it is unsurprising that the city has a large number of deaths (Kunasekaran, 2017).

Fourth, in terms of net migration, metropolitan cities (capital metropolitan and regional metropolitan areas) are the only settlements projected to experience net-inflows. In contrast, small towns/villages are projected to have the largest net-outflows, albeit at declining levels over time. However, within the capital metropolitan area, there will only be net overall in-migration because net migration into the suburbs offsets net-outflows from the core. For other settlement types, the results show no major changes in the future. In terms of age groups, young adult migrants are projected to have large net-inflows into capital metropolitan suburban areas. In contrast, small towns/villages are projected to have large net-outflows. Although this analysis does not include the net flows between settlement types, the contradictory net migration patterns of these settlement types indicate a classical rural-urban migration pattern. Apart from rural-urban migration, the influx of young adults in capital metropolitan suburban areas is also due to urban-urban migration. These results are similar to those in Chapter 5, with urban-urban migration more dominant than rural-urban migration in the capital metropolitan area by 2000, which should continue so since migration projections are assumed to be constant in the future.

Fifth, the population in all settlement types is projected to grow in size. However, capital metropolitan suburban areas are projected to have the largest population change in the future. In contrast, the future population in the capital metropolitan core is projected to diminish continuously. Apart from the capital metropolitan core, the share and growth in small towns/villages are also projected to decline continuously in the future. Regional metropolitan and intermediate-sized cities, on the other, will see no major changes in relative standing and will remain stable throughout the projection period.

Sixth, all settlements are projected to experience a major change in age structure in the future. The change of age structure between the starting and final years of projection indicates that future population growth will mainly be sustained by adults or the working-age population. This is due to the large proportion of the female population of reproductive age, resulting in more births despite a future decline in fertility below replacement level. This phenomenon is known as population or demographic momentum, which arises when previous high fertility results in a large proportion of the female population being of reproductive ages, hence leading to a large number of births (Keyfitz, 1971). The increase of the working-age population may also have a significant impact on economic growth which relates to the demographic dividend. Demographic dividend occurs when economic growth is accelerated by the decline of fertility and mortality and a change of the age structure (Gribble & Bremner, 2012). As time passes, the proportion of children grows smaller than the proportion of the working-age population, hence there are fewer children to support, leading to significant economic opportunities due to the implementation of policies and investments (Gribble & Bremner, 2012).

Seventh, rapid suburbanisation will continue if recent migration trends persist in the future. However, the long suburbanisation process in the capital metropolitan area (until 2040 at least) may worsen the current conditions of the city. Urban sprawl has been a major issue in Malaysia since the 1980s due to rapid suburbanisation in metropolitan cities caused by uncontrolled urban development and growth (Abdullah, 2012; Hasan & Nair, 2014). The increasing number of large-scale urban development projects such as the new inter-city rail project, East Coast Rail Link, announced in 2016 that will connect cities, including capital metropolitan suburban areas (Malaysia Rail Link, 2019), and continuing urban sprawl problems may have a major impact on urban development growth and the natural environment, pressuring the Malaysian government to provide more expenditure for housing, infrastructure, and amenities.

Finally, the projections capture similar trends as those postulated in demographic transition theory and mobility transition theory. Fertility and mortality levels in Malaysia have been continuously declining since 1980 and are projected to decline even further in the future. Because this chapter follows similar assumptions as the official projections, the total fertility rate is assumed to decline to below the replacement level by 2040. Further, life expectancy is assumed to increase for both males and females in the future. Furthermore, rural-urban migration has been significant in most settlement types in Malaysia (except in the capital metropolitan area) since 1980 and should remain so because the recent migration rates used for the projections in this thesis will remain unchanged in the future. Based on this evidence, if these assumptions are correct, Malaysia will see a similar demographic pattern as hypothesized in these theories, which is Phase III of demographic and mobility transition.

8.3 Study limitations

Although this thesis successfully addressed the aim and objectives, there are also inevitable issues and limitations that arise in the methodological approach.

First, differential urbanisation theory emphasises only migration change in determining the urbanisation transition process. However, urbanisation, polarisation reversal, and counterurbanisation are caused by not only migration (e.g., urban-rural, urban-urban, rural-urban, and rural-rural) but also natural increase of urban and rural populations. Therefore, it is important to integrate other theories such as mobility transition or demographic transition to address migration and demographic change in relation to the urbanisation process (Dyson, 2011).

Second, due to the limitation of fertility and mortality information by small-area units, this thesis does not include an examination of the natural causes of urbanisation from 1980-2010. In contrast, an examination of internal migration change is included in Chapter 5 since migration data is available for small-area units. Nevertheless, the importance of natural increase to population change and the overall population in Malaysia can be captured through the analysis of the relative importance of net migration in Chapter 5.

Third, it is hard to distinguish settlement types using district units. For example, the core city and suburban areas in regional metropolitan areas cannot be distinguished because the district covers both areas. As another example, small towns/villages cannot be separated because of the same issue. This makes it harder to identify which migration pattern is more dominant, rural-urban or urban-urban (small towns to larger towns). On balance, rural-urban migration possibly has a greater influence because settlements were established as villages in the beginning but then evolved into small towns as time passed. Settlement classification and boundaries made from *mukim* units are more accurate than those made from district units since they are smaller (Chapter 4). However, far more data are available at the district level. Therefore, more analyses were necessarily conducted using a classification of settlement types based on district-level data (Chapter 5 to 7).

Finally, despite the limitations of data and different methods used for the projections in this thesis, these projections match the official projections at the national level and are reasonably similar at the state level. All future assumptions on demographic inputs (fertility, mortality, and migration) of the projections in this thesis more or less mirror the assumptions made for official projections. A key difference is that the projections of this thesis exclude non-Malaysian citizens. A major question is what the implications of omitting them are. Apart from data limitations, non-Malaysian citizens accounted for only 8 percent from the overall population in 2010. By 2040, they is projected to increase to 9 percent. Clearly, the change is minimal because the Malaysian government strictly controls the future number of immigrants via work permission and visas. Further, because Malaysian citizens account for the majority of the population, projections for them should be sufficient for explaining the future population of Malaysia as a whole. Furthermore, if non-Malaysian citizens were considered in the projections of this thesis, the results would show the same story (e.g., more concentration in urban than rural areas) due to the strong government controls.

8.4 Future urbanisation pathway in Malaysia

In relation to differential urbanisation theory, Malaysia will remain in the APC stage until 2040 as a result of the dominance of positive net-inflows into capital metropolitan suburban areas. As time passes, the net-inflows into capital metropolitan suburban areas will, however, decrease, indicating a slowdown in the suburbanisation process. In contrast, the net-inflows in regional metropolitan areas will slowly increase over the years. This situation ties in to the differential urbanisation theory hypothesis, whereby Malaysia will start moving towards the polarisation reversal stage based on the shrinking dominance and attractiveness of the largest city (capital metropolitan suburban areas) and growing net-inflows in intermediate-sized cities (regional metropolitan areas). However, the difference of net-inflows between these settlements is relatively large, hence polarisation reversal is not expected to happen until shortly after 2040.

Although the projection results show the country has the potential to experience polarisation reversal in the future, the continuous rapid growth of urban development in the largest city, the capital metropolitan area, may distort the process (e.g., large-scale projects such as new townships, high-speed rail). It is true that the government has plans to balance the population across regions (e.g., through rural and regional settlement schemes and the establishment of educational institutions far from metropolitan cities to stimulate growth in other areas). On the other hand, (a) previous Malaysian economic policy interventions aimed at ethnic redistribution had only limited success and (b) as a liberal free-market economy, Malaysia is unlikely to have the political will required to impose policies that run entirely counter to the majority of its citizen's desires. For example, if people want to migrate into capital metropolitan suburban areas due to huge economic potential, nobody will stop them, with the only restriction their financial capability. Due to less government intervention, Malaysia in the future may look more like a standard developing country in terms of its urbanisation pathway. Obviously, there are many urban planning policies and laws, but so far they require tweaking at the edges to work.

In India, the change of the urbanisation pattern into polarisation reversal was mainly due to the effectiveness of various programmes and policies during the post-independence period to have a balanced settlement size and population growth. One of the policies limited the concentration in large cities by encouraging concentration in other cities through infrastructure development and the establishment of transportation networks (Mookherjee, 2003; Mookherjee & Geyer, 2011; Seto, 2011). However, the level of government intervention level is low, which is similar to Malaysia, where all programmes and policies imposed are meant to encourage the population to live in other cities or rural areas, not force them to do so. In China, the level of government intervention is stricter than in India and Malaysia since it is a communist country. For example, recent news in China highlighted that the ethnic minorities (especially Muslims) were forced to live in a designated area. More importantly, due to the inadequacy of physical infrastructure and the lack of institutional capacity to decentralise, this might lead India to re-urbanise instead of counterurbanise (Jain et al., 2013).

8.5 Recommendations for future research

There are three potential areas in which the research in this thesis could be extended in the future. First, more urban-rural classifications can be added in addition to just cities of different sizes (i.e., large, medium, and small), such as areas that are based on economic or social activities or geographic characteristics (e.g., industrial, agricultural, education, tourism, or coastal areas). There is a growing literature that examines population and migration change for different scales of cities and rural areas (see Champion et al., 1998; Dennett, 2010; Rees et al., 1996; Simpson & Finney, 2009). The recently published the second National Urbanisation Policy for Malaysia outlines the characteristics of each city, which could help determine the characteristics of cities or rural areas (Federal Department of Town and Country Planning in Peninsular Malaysia, 2016). Alternatively, but equivalently, districts could be designed on the basis of their geodemographic characteristics, as in the 2011 OLS district-level geodemographic classification.

Second, this thesis does not examine the natural causes of urbanisation from 1980-2010. This could be estimated for this period due to availability of fertility and mortality inputs. The estimates can be made by applying the cohort-component model. Although the fertility and mortality inputs are at the state level, similar methods can be used as in the projection chapter (Chapter 7) by assuming each district has similar inputs to the state of they are a part. There are several possible outcomes if this idea is implemented: 1) observe previous trends of natural increase by small-area units; 2) compare the contribution of natural increase and internal migration to urbanisation; and 3) test the validity of related theories (e.g., demographic and mobility transition theories).

Finally, the migration and projection models in this thesis (Chapter 6 and Chapter 7) could be revisited using the full data available from the Malaysia Department of Statistics, rather than just what is publicly available. In this sense, this thesis provides a prototype for a potential future official statistical work programme.

8.6 Concluding remark

To date, few urbanisation studies in Malaysia have adopted small-area units and modelling techniques for analysis**.** The lack of research on this matter inspired the author to conduct such work and complete this thesis. Only a few studies have applied modelling techniques to examine migration. For example, studies have examined the relationship of migration to fertility (Bach, 1981), career (Chattopadhyay, 1998), ethnic concentration (Chitose, 2001) and income and unemployment (Hussain & Abdullah, 2014). This thesis has successfully explored recent and future population growth, internal migration, and urbanisation in Malaysia by small-area units, mainly through the implementation of differential urbanisation theory. It sheds new light on historic and potential future trends, by settlement type and by district. Nevertheless, there remains lots of room for improvement and potential research that can be done in the future. Even so, it is hoped that all findings in this thesis will benefit the Malaysian government and those who are interested in this topic.

Bibliography

Abdullah, J. (2003). The Suburbanisation of the Kuala Lumpur Metropolitan Region. *Journal of the Malaysian Institute of Planners*.

Abdullah, J. (2012). City Competitiveness and Urban Sprawl: Their Implications to Socio-Economic and Cultural Life in Malaysian Cities. *Procedia - Social and Behavioral Sciences*, *50*(July), 20–29. https://doi.org/10.1016/j.sbspro.2012.08.012

Abdullah, J., Yahaya, M. Z., Mohd Yunus, M. Z., & Safudin, M. S. M. A. (2009). Urban Sprawl in Malaysia : Evidences From Three Largest Metro Polit an Areas. *Journal of the Malaysian Institute of Planners*, *7*, 69–82.

Aiken, S. R., & Leigh, C. H. (1975). Malaysia'S Emerging Conurbation. *Annals of The Association Of American Geographers*, *65*(4), 546–562.

Andreev, Kirill; Kantorová, Vladimíra; Bongaarts, J. (2013). Demographic Components of Future Population Growth. *UN Technical Paper Series*, (3), 26. https://doi.org/10.4135/9781412952446.n605

Andrew, G., & Jennifer, H. (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.

Argent, N., & Rolley, F. (2008). Counterurbanisation in Australia: Two Case Studies. *Geodate*, 4–7. https://doi.org/10.5817/BSE2012-2-9

Armstrong, J. S. (1985). *Long-Range Forecasting*. *John Wiley*. New York.

Awumbila, M. (2017). *Drivers of Migration and Urbanization in Africa: Key Trends and Issues*. Legon, Ghana.

Bach, R. L. (1981). Migration and Fertility in Malaysia: A Tale of Two Hypotheses. *The International Migration Review*, *15*(3), 502–521.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bell, M., Charles-Edwards, E., Ueffing, P., Stillwell, J., Kupiszewski, M., & Kupiszewska, D. (2015). Internal Migration and Development: Comparing Migration Intensities Around the World. *Population and Development Review*, *41*(1), 33–58. https://doi.org/10.1111/j.1728-4457.2015.00025.x

Bell, M., & Muhidin, S. (2009). *Cross-National Comparison of Internal Migration*. *Human Development Research Paper 2009/30*. New York.

Bernard, A., Bell, M., & Edwards, E. C. (2014). Life-course transitions and the age profile of internal migration. *Population and Development Review*, *40*(2), 213–239. https://doi.org/10.1111/j.1728-4457.2014.00671.x

Berry, B. J. L. (1980). Urbanization and Counterurbanization in the United States. *The Annals of the American Academy of Political and Social Science, Changing Cities: A Challenge to Planning*, *451*, 13–20.

Blue, L., & Espenshade, T. J. (2011). Population momentum across the demographic transition. *Population and Development Review*, *37*(4), 721–747.

Bodvarsson, Ö. B., & Van den Berg, H. (2013). *The Economics of immigration: Theory and policy, 2nd ed. 2013*. *The Economics of Immigration: Theory and Policy, 2nd Ed. 2013*. https://doi.org/10.1007/978-1-4614-2116-0

Bonifazi, C., & Heins, F. (2003). Testing the differential urbanisation model for Italy. *Tijdschrift Voor Economische En Sociale Geografie*, *94*(1), 23–37. https://doi.org/10.1111/1467-9663.00234

Borjas, G. J. (1994). *The Economics of Immigration*. *Journal of Economic Literature* (Vol. 32). https://doi.org/10.2307/2728791

Burch, K. T. (2018). *Model-Based Demography*.

Champion, T. (2003). Testing the differential urbanisation model in Great Britain, 1901-91. *Journal of Economic and Social Geography*, *94*(1).

Champion, T. (2005). The counterurbanisation cascade in England and Wales since 1991: The evidence of a new migration dataset. *Belgeo*, (1–2), 85–101. https://doi.org/10.4000/belgeo.12440

Champion, T., Fotheringham, S., Rees, P., Boyle, P., & Stillwell, J. (1998). *The Determinants of Migration Flows in England: A Review of Existing Data and Evidence*.

Chan, K. W. (2012). Internal labor migration in China: Trends, geography and policies. *Population Distribution, Urbanization, Internal Migration and Development: An International Perspective*, 81–102.

Chan, M. F., & Kamala Devi, M. (2015). Factors affecting life expectancy: Evidence from 1980-2009 data in Singapore, Malaysia, and Thailand. *Asia-Pacific Journal of Public Health*, *27*(2), 136–146. https://doi.org/10.1177/1010539512454163

Chattopadhyay, A. (1998). Gender, Migration, and Career Trajectories in Malaysia. *Demography*, *35*(3), 335–344.

Chitose, Y. (2001). The effects of ethnic concentration on internal migration in Peninsular Malaysia. *Asian and Pacific Migration Journal*, *10*(2), 241–272. https://doi.org/10.1177/011719680101000202

Coombes, M., Longa, R. D., & Raybould, S. (1989). Counterubanisation in Britain and Italy : A Comparative Critique of the Concept , Causation and Evidence. *Progress in Planning*, *32*, 1–70.

Dark, S. J., & Bram, D. (2007). The modifiable areal unit problem (MAUP) in physical geography. *Progress in Physical Geography*, *31*(5), 471–479. https://doi.org/10.1177/0309133307083294

Del Carpio, X., Özden, Ç., Testaverde, M., & Wagner, M. (2015). Local labor supply responses to immigration. *Scandinavian Journal of Economics*, *117*(2), 493–521. https://doi.org/10.1111/sjoe.12099

Deming, W. E., & Stephan, F. F. (1940). On a Least Squares Adjustment of a Sampled Frequency Table When the Expected Marginal Totals are Known. *The Annals of Mathematical Statistics*, *11*(4), 427–444. https://doi.org/10.1214/aoms/1177705148

Dennett, A. R. (2010). *Understanding Internal Migration in Britain at the Start of the 21st Century by*. University of Leeds. https://doi.org/10.1002/psp.554

Dennett, A., & Wilson, A. (2013). A multilevel spatial interaction modelling framework for estimating interregional migration in Europe. *Environment and Planning A*, *45*(6), 1491–1507. https://doi.org/10.1068/a45398

Department of Statistics Malaysia. (2018). *Statistics on Causes of Death Malaysia 2018*.

Dewan Bandaraya Kuala Lumpur. (2003). Kuala Lumpur Structure Plan 2020 :

Economic Base and Population. Retrieved June 4, 2018, from http://www.dbkl.gov.my/pskl2020/english/economic_base_and_population/index.htm

Dormann, C. F. (2016). Overdispersion, and how to deal with it in R and JAGS. Retrieved August 5, 2018, from https://github.com/biometry/APES/blob/master/LectureNotes/2016-JAGS/Overdispersion/OverdispersionJAGS.md

Dyson, T. (2011). The role of the demographic transition in the process of urganization. *Population and Development Review*, *37*(Supplement), 34–54.

Ernestina, C. (2002). *Population trends in developing countries [online]. London: LSE Research Online*. Retrieved from http://eprints.lse.ac.uk/268/

Faizah, M. T. (2007). The 40th Session Of The Commission On Population And Development On Agenda Item 4: General Debate On National Experience In Population Matters: The Changing Age Structures Of Populations And Their Implications For Development.

Federal Department of Town and Country Planning Peninsular Malaysia. (2016). *The Second National Urbanisation Policy*.

Field, A. (2009). *Discovering Statistics Using SPSS* (Third Edit). Sage Publications.

Fielding, A. J. (1982). Counterurbanisation in Western Europe. *Progress in Planning*, *17*, 1–52.

Fox, J., & Weisberg, S. (2011). *An R Companion to Applied Regression* (Second edi). Thousand Oaks CA: Sage Publications.

Fox, S. (2012). Urbanization as a Global Historical Process : Theory and Evidence from sub-Saharan Africa. *Population and Development Review*, *38*(2), 285–310.

Free Malaysia Today. (2017). Poor health among Malaysians due to work stress, says survey. Retrieved April 23, 2019, from https://www.freemalaysiatoday.com/category/nation/2017/11/18/poor-health-among-msians-due-to-work-stress-says-survey/

Gallup, J. L. (1996). Migration in Malaysia: Heterogeneity and Persistence.

Gartner, A., Farewell, D., Dunstan, F., & Gordon, E. (2008). *Differences in Mortality Between Rural and Urban Areas in England and Wales, 2002-04. Health Statistics Quarterly* (Vol. Autumn).

Gedik, A. (2003). Differential urbanization in Turkey: 1955-2000. In *43rd Congress of The European Regional Science Association* (pp. 1–17).

Gelman, A., & Su, Y.-S. (2018). *R package version 1.10-1*. Retrieved from https://cran.r-project.org/package=arm

Geyer, H. S. (2003). Differential urbanisation in South Africa - a further exploration. *Tijdschrift Voor Economische En Sociale Geografie*, *94*(1), 89–99. https://doi.org/10.1111/1467-9663.00239

Geyer, H. S., & Kontuly, T. (1993). A Theoretical Foundation for the Concept of Differential Urbanization. *International Regional Science Review*, *15*(2), 157–177. https://doi.org/10.1177/016001769301500202

Geyer Jr, H. S., & Geyer, H. S. (2015). Disaggregated Population Migration Trends in South Africa Between 1996 and 2011: a Differential Urbanisation Approach. *Urban Forum*, *26*, 1–13. https://doi.org/10.1007/s12132-014-9229-1

Geyer Jr, H. S., & Geyer, H. S. (2016). Polarisation reversal in South Africa: how widespread is the trend? *South African Geographical Journal*, *98*(2), 289–307. https://doi.org/10.1080/03736245.2015.1028986

Geyer Jr, H. S., Geyer, H. S., Plessis, D. J. du, & Eeden, A. van. (2012). Differential urbanisation trends in South Africa-regional and local equivalents. *Environment and Planning A*, *44*(12), 2940–2956. https://doi.org/10.1068/a4528

Geyer, N., & Geyer, H. (2017). Counterurbanisation: South Africa in wider context. *Environment and Planning A*, *49*(7), 1575–1593. https://doi.org/10.1177/0308518X17701728

Geyer, N. P. (2018). Counterurbanization in South Africa : Measuring migration significance, (November 2017), 3–14. https://doi.org/10.1111/rsp3.12105

Gottmann, J. (2004). Urbanisation and Employment, (December 1977).

Greenwood, M. (1997). *Internal migration in developed countries. In M. Rosenzweig & O. Stark (Eds.). Handbook of population and family economics* (Mark R. Ro). Amsterdam: Elsevier.

Gribble, J., & Bremner, J. (2012). The challenge of attaining the demographic dividend. *Population Reference Bureau*, (September), 1–6. Retrieved from http://www.prb.org/pdf12/demographic-dividend.pdf

Gwebu, T. D. (2006). Towards a theoretical exploration of the differential urbanisation model in Sub-Saraharan Africa: The Botswana case. *Tijdschrift Voor Economische En Sociale Geografie*, *97*(4), 418–433. https://doi.org/10.1111/j.1467-9663.2006.00346.x

Halliday, J., & Coombes, M. (1995). In search of counterurbanisation: Some evidence from Devon on the relationship between patterns of migration and motivation. *Journal of Rural Studies*, *11*(4), 433–446. https://doi.org/10.1016/0743-0167(95)00032-1

Hansell, C. R., & Clark, W. A. V. (1970). The Expansion of the Negro Ghetto in Milwaukee a Description and Simulation Model. *Tijdschrift Voor Economische En Sociale Geografie*, *61*(2), 267–277. https://doi.org/10.1111/j.1467-9663.1970.tb00491.x

Harris, J. R., & Todaro, M. P. (1970). Migration, Unemployment and Development: A Two-Sector Analysis. *The American Economic Review*, *60*(1), 126–142. https://doi.org/10.1016/0165-1765(84)90160-5

Hasan, A. R., & Nair, P. L. (2014). Urbanisation and growth of metropolitan centres in Malaysia. *Malaysian Journal of Economic Studies*, *51*(1), 87–101.

Heikkila, E. (2003). Differential urbanisation in Finland. *Tijdschrift Voor Economische En Sociale Geografie*, *94*(1), 49–63. https://doi.org/10.1111/1467-9663.00236

Henderson, V. (2002). Urbanization in Developing Countries. *The World Bank Research Observer*, *17*(1), 89–112. https://doi.org/10.2307/3986401

Hirschman, C. (1976). Recent Urbanization Trends in Peninsular Malaysia. *Demography*, *13*(4), 445–461.

Hirschman, C. (1980). Demographic Trends in Peninsular Malaysia, 1947-75. *Population and Development Review*, *6*(1), 103–125.

Hosszú, S. (2009). *Counterurbanization A literature study*. University of Southern Denmark, Esbjerg.

Hugo, G. (2002). Changing patterns of population distribution in Australia. *Journal of Population Research and NZ Population Review*, *Joint Spec*(September), 1–21.

Hussain, N. E., & Abdullah, N. (2014). The Relationship between Rural-Urban Migration , Household Income and Unemployment : Malaysia Case Study. *International Journal of Managerial Studies and Research (IJMSR)*, *2*(8), 17–24.

IBM Corp. (2016). *Version 24.0*. Armonk, NY.

J, P., D, B., S, D., D, S., & R Core Team. (2016). *R package version 3.1-128*. Retrieved from http://cran.r-project.org/package=nlme

Jain, M., Siedentop, S., Taubenbock, H., & Namperumal, S. (2013). From Suburbanization to Counterurbanization?: Investigating Urban Dynamics in the National Capital Region Delhi, India. *Environment and Urbanization Asia*, *4*(2), 247–266. https://doi.org/10.1177/0975425313510765

Jali, M. R. M. (2009). *Internal Migration in Malaysia : Spatial and Temporal Analysis*. School of Geography, University of Leeds.

Jedwab, R., Christianesen, L., & Gindelsky, M. (2014). *Demography, Urbanization and Development: Rural Push, Urban Pull and Urban Push?*

Joseph, G. (1975). A Markov Anaylsis of Age/Sex Differences in Inter-regional Migration in Great Britain. *Regional Studies*, *9*(1), 69–78. https://doi.org/10.1080/09595237500185061

Kanaroglou, P. S., & Braun, G. O. (1992). The Pattern of Counterurbanization in the Federal-Republic-of-Germany, 1977-85. *Environment and Planning A*, *24*(4), 481–496.

Kasarda, J. D., & Crenshaw, E. M. (1991). Third World Urbanization : Dimensions, Theories, And Determinants. *Annual Review Sociology*, *17*, 467–501.

Keyfitz, N. (1971). On the Momentum of Population Growth. *Demography*, *8*(1), 71–80.

Khan, A. (2008). The Industrial Revolution and the Demographic Transition. *Business Review*, 9–15.

Kim, J. H., & Han, J. H. (2014). Myths of migration on retirement in Korea: Do the elderly move toless dense areas? *Habitat International*, *41*, 195–204. https://doi.org/10.1016/j.habitatint.2013.08.003

King, A., Walker, A., & Bækgaard, H. (2002). *Modelling Overseas Migration in Dynamod-2. NETSEM Working Paper*.

Kirk, D. (1996). Demographic Transition Theory. *Population Studies*, *50*(3), 361–387.

Kontuly, T., & Vogelsang, R. (1988). Explanations for the intensification of counterurbanization in the federal republic of germany. *Professional Geographer*, *40*(1), 42–54. https://doi.org/10.1111/j.0033-0124.1988.00042.x

Kontuly, T., Wiard, S., & Vogelsang, R. (1986). Counterurbanization in The Federal Republic of Germany. *The Professional Geographer*, *38*(2), 170–181. https://doi.org/10.1111/j.0033-0124.1986.00170.x

Kulu, H., & Boyle, P. J. (2009). High fertility in city suburbs: Compositional or contextual effects? *European Journal of Population*, *25*(2), 157–174. https://doi.org/10.1007/s10680-008-9163-9

Kulu, Hill. (2005). Migration and fertility: Competing hypotheses re-examined. *European Journal of Population*, *21*(1), 51–87. https://doi.org/10.1007/s10680-005-3581-8

Kulu, Hill. (2013). Why Do Fertility Levels Vary between Urban and Rural Areas? *Regional Studies*, *47*(6), 895–912. https://doi.org/10.1080/00343404.2011.581276

Kulu, Hill, Boyle, P. J., & Andersson, G. (2009). High suburban fertility: Evidence from four Northern European countries. *Demographic Research*, *21*, 915–944. https://doi.org/10.4054/DemRes.2009.21.31

Kulu, Hill, & Vikat, A. (2007). Fertility differences by housing type: The effect of housing conditions or of selective moves? *Demographic Research*, *17*, 775–802. https://doi.org/10.4054/DemRes.2007.17.26

Kulu, Hill, & Washbrook, E. (2014). Residential context, migration and fertility in a modern urban society. *Advances in Life Course Research*, *21*, 168–182. https://doi.org/10.1016/j.alcr.2014.01.001

Kunasekaran, T. (2017). *Death In Malaysia- What are the STatistic telling us? Khazanah Research Institute*.

Kurek, S., Wójtowicz, M., & Gałka, J. (2015). The changing role of migration and natural increase in suburban population growth: The case of a non-capital post-socialist city (the Krakow metropolitan area, Poland)? *Moravian Geographical Reports*, *23*(4), 59–70. https://doi.org/10.1515/mgr-2015-0025

Lall, S. V., Selod, H., & Shalizi, Z. (2006). *Rural-urban migration in developing countries : a survey of theoretical predictions and empirical findings*. *World Bank Policy Research Working Paper Series*. https://doi.org/http://dx.doi.org/10.1596/1813-9450-3915

Lee, H., Stuart R., G., & Robert, E. (2011). Projecting a Gridded Population of the World Using Ratio Methods of Trend Extrapolation. *Human Population: Its Influences on Biological Diversity*, *214*(1989), 103–124. https://doi.org/10.1007/978-3-642-16707-2

Lestari. (1997). *Urbanization and Environment in Malaysia: Managing the Impact*.

Long, L. (1988). *Migration and residential mobility in the United States*. New York: Russell Sage Foundation.

Lotfi, S. (1998). *Development and Urbanisation: The Case Study of Amol and Mazandaran Province, Iran*. University of Glassgow.

Mahari, Z. (2011). Demographic Transition in Malaysia : The Changing Roles of Women.

Malaysia Rail Link. (2019). Overview of East Coast Rail Link (ECRL). Retrieved April 19, 2019, from http://www.mrl.com.my/en/ecrl/overview/

Masron, T., Yaakob, U., Mohd Ayob, N., & Mokhtar, A. S. (2012). Population and spatial distribution of urbanisation in Peninsular. *Malaysia Journal of Society and Space*, *2*(2), 20–29.

Mcgranahan, G., & Satterthwaite, D. (2014). *Urbanisation concepts and trends*. *International Institute for Environment and Development*.

Mihermutu, B. A. (2011). *Causes and Consequences of Rural-Urban Migration: the Case of Woldiya Town, North Ethiopia*. University of South Africa. https://doi.org/10.1017/CBO9781107415324.004

Mitchell, C. J. A. (2004). Making sense of counterurbanization. *Journal of Rural Studies*, *20*(1), 15–34. https://doi.org/10.1016/S0743-0167(03)00031-7

Mookherjee, D. (2003). Differential urbanisation model: the case of a developing country, India 1961-91. *Tijdschrift Voor Economische En Sociale Geografie*, *94*(1), 38–48. https://doi.org/10.1111/1467-9663.00235

Mookherjee, D., & Geyer, M. H. S. (2011). Urban Growth in the National Capital Region of India: Testing the Differential Urbanisation Model. *Tijdschrift Voor Economische En Sociale Geografie*, *102*(1), 88–99. https://doi.org/10.1111/j.1467-9663.2010.00599.x

Morrill, R. L. (1965). The Negro Ghetto: Problems and Alternatives. *Geographical Review*, *55*(3), 339–361.

Moy, E., Garcia, M. C., Bastian, B., Rossen, L. M., Ingram, D. D., Faul, M., … Iademarco, M. F. (2017). Leading Causes of Death in Nonmetropolitan and Metropolitan Areas— United States, 1999–2014. *MMWR. Surveillance Summaries*, *66*(1), 1–8. https://doi.org/10.15585/mmwr.ss6601a1

Nefedova, T., & Treivish, A. (2003). Differential Urbanisation in Russia. *Royal*

Dutch Geographical Society KNAG, *94*(1), 75–88.
Newell, C. (1990). Methods and Models in Demography. *Methods and Models in Demography*.
O'Reilly, G., O'Reilly, D., Rosato, M., & Connolly, S. (2007). Urban and rural variations in morbidity and mortality in Northern Ireland. *BMC Public Health*, *7*, 1–6. https://doi.org/10.1186/1471-2458-7-123
Osman, S., Abdullah, J., & Nawawi, A. H. (2017). The financial costs of urban sprawl: Case study of Penang State. *Planning Malaysia*, *15*(6), 13–24.
Osman, S. bt, Nawawi, A. H., & Abdullah, J. (2009). Urban Sprawl and Its Financial Cost: - A Conceptual Framework. *Asian Social Science*, *4*(10), 39. https://doi.org/10.5539/ass.v4n10p39
Peng, X., Chen, X., & Cheng, Y. (2011). Urbanization And Its Consequences. In *Encyclopedia of Life Support Systems (EOLSS)* (Vol. II).
R Core Team. (2016). *R package version 0.8-67*. Retrieved from https://cran.r-project.org/package=foreign
Ravenstein, E. G. (1885). The Laws of Migration. *Journal of the Statistical Society of London*, *48*(2), 167–235.
Rayer, S. (2015). Demographic Techniques: Small-area Estimates and Projections. *International Encyclopedia of the Social & Behavioral Sciences: Second Edition*, *5*, 162–169. https://doi.org/10.1016/B978-0-08-097086-8.31015-7
Rees, P., Durham, H., & Kupiszewski, M. (1996). *Internal Migration and Regional Population Dynamics in Europe: United Kingdom Case Study*. https://doi.org/10.1017/CBO9781107415324.004
Reher, D. S. (2004). The demographic transition revisited as a global process. *Population, Space and Place*, *10*(1), 19–41. https://doi.org/10.1002/psp.313
Richardson, H. W. (1980). Polarisation Reversal in Developing Countries. *Papers in Regional Science*, *45*, 67–85.
Rii, H. U., & Ahn, J.-S. (2002). *Urbanization and its Impact on Seoul, Korea. [Taipei] : Chung-hua Institution for Economic Research*.
Rodriguez, J., & Rowe, F. (2018). Internal migration and spatial de-concentration of population in Latin America. *International Union for the Scientific Studies of Population*, (March). Retrieved from https://www.cepal.org/celade/migracion/migracion_interna/
Rodríguez, J., & Rowe, F. (2018). How is internal migration reshaping metropolitan populations in Latin America? A new method and new evidence. *Population Studies*, *72*(June), 253–273. https://doi.org/10.1080/00324728.2017.1416155
Rogers, A. (1966). A markovian analysis of migration differentials. In *American Statistical Association* (pp. 452–466).
Rogers, A. (1990). Requiem for the Net Migrant, *22*(4), 283–300.
RStudio Team. (2016). RStudio: Integrated Development for R. RStudio, Inc. Boston, MA. Retrieved from http://www.rstudio.com/
Sapiro, P. (2017). Explaining geographic patterns of small group internal migration. *Population, Space and Place*, *23*(8), 1–11. https://doi.org/10.1002/psp.2078
Saracoglu, S., & Roe, T. L. (2004). *Rural-UrbanMigration and Economic Growth in Developing Countries*. *Society for Economic Dynamics, 2004 Meeting Papers*.
Sendut, H. (1966). Urbanization in Malaysia. *Asian Survey*, *6*(9), 484–491.
Seto, K. C. (2011). Exploring the dynamics of migration to mega-delta cities in Asia and Africa: Contemporary drivers and future scenarios. *Global Environmental Change*, *21*(SUPPL. 1), S94–S107. https://doi.org/10.1016/j.gloenvcha.2011.08.005

Shamshad. (2012). Rural to Urban Migration: Remedies to Control. *Golden Research Thoughts*, *2*(4), 1–6.
Shuid, S. (2004). *Urbanization and Housing in Kuala Lumpur City Centre: Issues and Future Challenges*. Retrieved from https://books.google.co.uk/books?id=cTJTAQAACAAJ
Simpson, L., & Finney, N. (2009). Spatial patterns of internal migration: Evidence for ethnic groups in Britain. *Population, Space and Place*, *15*(1), 37–56. https://doi.org/10.1002/psp.497
Smith, S. K., Jeff, T., & David A., S. (2017). *A Practitioner's Guide to State and Local Population Projections*. *Canadian Studies in Population* (Vol. 42). https://doi.org/10.25336/p61883
Soon, J. J. (2010). When do students intend to return? Determinants of students' return intentions using a multinomial logit model. *International Journal of Business and Society*, *11*(1), 17–32.
Stillwell, J., & Congdon, P. (1991). *Migration models: macro and micro approaches*. Belhaven Press.
Suryawanshi, D., Sharma, V., Saggurti, N., & Bharat, S. (2016). Factors associated with the likelihood of further movement among mobile female sex workers in India: A multinomial Logit approach. *Journal of Biosocial Science*, *48*(4), 539–556. https://doi.org/10.1017/S0021932015000267
Tacoli, C., & Satterthwaite, G. M. D. (2015). *Urbanization, Rural–urban Migration and Urban Poverty*. *Migration Report*.
Tammaru, T. (2003). Urban and rural population change in Estonia: patterns of differentiated and undifferentiated urbanisation. *Tijdschrift Voor Economische En Sociale Geografie*, *94*(1), 112–123. https://doi.org/10.1111/1467-9663.00241
Tammaru, T., Kulu, H., & Kask, I. (2004). Urbanization, Suburbanization, and Counterurbanization in Estonia. *Eurasian Geography and Economics*, *45*(3), 212–229. https://doi.org/10.2747/1538-7216.45.3.212
Tobler, W. R. (1970). A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, *46*(2), 234–240.
UCLA: Statistical Consulting Group. (n.d.). Negative Binomial Regression | R Data Analysis Examples. Retrieved August 5, 2018, from https://stats.idre.ucla.edu/r/dae/negative-binomial-regression/
UN-Habitat. (2016). *Urbanization and Development: Emerging Futures*.
UNESCO, UNDP, IOM, & UN-Habitat. (2012). *Overview of Internal Migration in Malaysia*.
UNICEF. (2014). Malaysia Migration Profiles. Retrieved from https://esa.un.org/miggmgprofiles/indicators/indicators.HTM#asia
Vartiainen, P. (1989). The end of drastic depopulation in rural Finland: Evidence of counterurbanisation? *Journal of Rural Studies*, *5*(2), 123–136. https://doi.org/10.1016/0743-0167(89)90041-7
Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S.* (Fourth Edi). New York: Springer.
Vobruba, T., Körner, A., & Breitenecker, F. (2016). Modelling, Analysis and Simulation of a Spatial Interaction Model. *IFAC-PapersOnLine*, *49*, 221–225. https://doi.org/10.1016/j.ifacol.2016.11.054
W. Olowa, O. (2012). Determinants of migration and remittances in rural Nigeria. *Journal of Development and Agricultural Economics*, *4*(7), 191–198. https://doi.org/10.5897/jdae11.053

White, M. J. (2016). International Handbook of Migration and Population Distribution, 505–524. https://doi.org/10.1007/978-94-017-7282-2
Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Verlag New York: Springer. Retrieved from http://ggplot2.org
Wilson, A. G. (2006). A Family of Spatial Interaction Models, and Associated Developments. *Environment and Planning A: Economy and Space*, *3*(1), 1–32. https://doi.org/10.1068/a030001
Wilson, C. (2011). Understanding global demographic convergence since 1950. *Population and Development Review*, *37*(2), 375–388. https://doi.org/10.1111/j.1728-4457.2011.00415.x
Wilson, T. (2011). *A Review of Sub-Regional Population Projection Methods*.
Wong, D. W. S. (2004). *The Modifiable Areal Unit Problem (MAUP)*. In: Janelle D.G. Warf B. Hansen K. (eds) WorldMinds: Geographical Perspectives on 100 Problems, Springer, Dordrecht.
Woods, R. (2003). Urban-Rural Mortality Differentials: An Unresolved Debate. *Population and Development Review*, *29*(1), 29–46.
Woods, R. I. (1981). Spatiotemporal models of ethnic segregation and their implications for housing policy. *Environment and Planning A*, *13*(11), 1415–1433. https://doi.org/10.1068/a131415
Yaakob, U., Masron, T., & Masami, F. (2010). Ninety Years of Urbanization in Malaysia : A Geographical Investigation of Its Trends and Characteristics. *Journal of Ritsumeikan Social Sciences and Humanities*, *4*(3), 79–101.
Young, M. L. (2004). *Circuits of migration: a structural analysis of migration in Peninsular Malaysia*. Australian National University.
Založnik, M. (2011). *Iterative Proportional Fitting Theoretical Synthesis and Practical Limitations*. https://doi.org/10.13140/2.1.2480.9923
Zelinsky, W. (1971). Hypothesis of the Mobility Transition. *Geographical Review*, *61*(2), 219–249.
Zimmermann, K. F., & Constant, A. F. (2012). The Dynamics of Repeat Migration: A Markov Chain Analysis. *International Migration Review*, *46*(2), 362–388. https://doi.org/10.1111/j.1747-7379.2012.00890.x
Zipf, G. (1946). The P1 P2 / D hypothesis: On the intercity movement of persons. *American Sociological Review*, *11*(6), 677–686.

Appendices

A1: List of Districts that are classified into settlement type

State	District	Types of settlement
F.T. Kuala Lumpur	*F.T. Kuala Lumpur*	Capital metropolitan (core)
F.T. Putrajaya	*F.T. Putrajaya*	Capital metropolitan (suburb)
Selangor	*Gombak*	Capital metropolitan (suburb)
Selangor	*Klang*	Capital metropolitan (suburb)
Selangor	*Kuala Langat*	Capital metropolitan (suburb)
Selangor	*Kuala Selangor*	Capital metropolitan (suburb)
Selangor	*Petaling*	Capital metropolitan (suburb)
Selangor	*Sepang*	Capital metropolitan (suburb)
Selangor	*Ulu Langat*	Capital metropolitan (suburb)
Selangor	*Ulu Selangor*	Capital metropolitan (suburb)
Johor	*Johor Bahru*	Regional metropolitan
Johor	*Kulaijaya*	Regional metropolitan
Kedah	*Kuala Muda*	Regional metropolitan
Kedah	*Kulim*	Regional metropolitan
Pahang	*Kuantan*	Regional metropolitan
Pulau Pinang	*Barat Daya*	Regional metropolitan
Pulau Pinang	*S.P. Tengah*	Regional metropolitan
Pulau Pinang	*S.P. Utara*	Regional metropolitan
Pulau Pinang	*S.P.Selatan*	Regional metropolitan
Pulau Pinang	*Timur Laut*	Regional metropolitan
Terengganu	*Kemaman*	Regional metropolitan
F.T. Labuan	*F.T. Labuan*	Intermediate cities
Kedah	*Kota Setar*	Intermediate cities
Kedah	*Pokok Sena*	Intermediate cities
Kelantan	*Kota Bharu*	Intermediate cities
Kelantan	*Tumpat*	Intermediate cities
Melaka	*Melaka Tengah*	Intermediate cities
Negeri Sembilan	*Seremban*	Intermediate cities
Perak	*Kampar*	Intermediate cities
Perak	*Kinta*	Intermediate cities
Perlis	*Perlis*	Intermediate cities

Sabah	*Kota Kinabalu*	Intermediate cities
Sabah	*Penampang*	Intermediate cities
Sabah	*Putatan*	Intermediate cities
Sabah	*Tuaran*	Intermediate cities
Sarawak	*Asajaya*	Intermediate cities
Sarawak	*Bau*	Intermediate cities
Sarawak	*Kuching*	Intermediate cities
Sarawak	*Samarahan*	Intermediate cities
Terengganu	*Kuala Terengganu*	Intermediate cities
Johor	*Batu Pahat*	Small towns/ villages
Johor	*Kluang*	Small towns/ villages
Johor	*Kota Tinggi*	Small towns/ villages
Johor	*Muar*	Small towns/ villages
Johor	*Pontian*	Small towns/ villages
Johor	*Segamat*	Small towns/ villages
Kedah	*Baling*	Small towns/ villages
Kedah	*Bandar Baharu*	Small towns/ villages
Kedah	*Kubang Pasu*	Small towns/ villages
Kedah	*Langkawi*	Small towns/ villages
Kedah	*Pendang*	Small towns/ villages
Kedah	*Yan*	Small towns/ villages
Kelantan	*Bachok*	Small towns/ villages
Kelantan	*Machang*	Small towns/ villages
Kelantan	*Pasir Mas*	Small towns/ villages
Kelantan	*Pasir Puteh*	Small towns/ villages
Kelantan	*Tanah Merah*	Small towns/ villages
Melaka	*Alor Gajah*	Small towns/ villages
Melaka	*Jasin*	Small towns/ villages
Negeri Sembilan	*Jempol*	Small towns/ villages
Negeri Sembilan	*Kuala Pilah*	Small towns/ villages
Negeri Sembilan	*Port Dickson*	Small towns/ villages
Negeri Sembilan	*Rembau*	Small towns/ villages
Negeri Sembilan	*Tampin*	Small towns/ villages
Pahang	*Bentong*	Small towns/ villages
Pahang	*Cameron Highland*	Small towns/ villages
Pahang	*Maran*	Small towns/ villages
Pahang	*Temerloh*	Small towns/ villages
Perak	*Batang Padang*	Small towns/ villages
Perak	*Hilir Perak*	Small towns/ villages
Perak	*Kerian*	Small towns/ villages
Perak	*Kuala Kangsar*	Small towns/ villages

Perak	*Larut dan Matang*	Small towns/ villages
Perak	*Manjung (Dinding)*	Small towns/ villages
Perak	*Perak Tengah*	Small towns/ villages
Sabah	*Kota Belud*	Small towns/ villages
Sabah	*Kudat*	Small towns/ villages
Sabah	*Papar*	Small towns/ villages
Sabah	*Sandakan*	Small towns/ villages
Sabah	*Semporna*	Small towns/ villages
Sabah	*Tawau*	Small towns/ villages
Sarawak	*Matu*	Small towns/ villages
Sarawak	*Miri*	Small towns/ villages
Sarawak	*Sarikei*	Small towns/ villages
Sarawak	*Sibu*	Small towns/ villages
Selangor	*Sabak Bernam*	Small towns/ villages
Terengganu	*Besut*	Small towns/ villages
Terengganu	*Dungun*	Small towns/ villages
Terengganu	*Marang*	Small towns/ villages
Johor	*Ledang*	Remote villages
Johor	*Mersing*	Remote villages
Kedah	*Padang Terap*	Remote villages
Kedah	*Sik*	Remote villages
Kelantan	*Gua Musang*	Remote villages
Kelantan	*Jeli*	Remote villages
Kelantan	*Kuala Krai*	Remote villages
Negeri Sembilan	*Jelebu*	Remote villages
Pahang	*Bera*	Remote villages
Pahang	*Jerantut*	Remote villages
Pahang	*Lipis*	Remote villages
Pahang	*Pekan*	Remote villages
Pahang	*Raub*	Remote villages
Pahang	*Rompin*	Remote villages
Perak	*Ulu Perak*	Remote villages
Sabah	*Beaufort*	Remote villages
Sabah	*Beluran*	Remote villages
Sabah	*Keningau*	Remote villages
Sabah	*Kinabatangan*	Remote villages
Sabah	*Kota Marudu*	Remote villages
Sabah	*Kuala Penyu*	Remote villages
Sabah	*Kunak*	Remote villages
Sabah	*Lahad Datu*	Remote villages
Sabah	*Nabawan*	Remote villages

Sabah	*Pitas*	Remote villages
Sabah	*Ranau*	Remote villages
Sabah	*Sipitang*	Remote villages
Sabah	*Tambunan*	Remote villages
Sabah	*Tenom*	Remote villages
Sabah	*Tongod*	Remote villages
Sarawak	*Belaga*	Remote villages
Sarawak	*Betong*	Remote villages
Sarawak	*Bintulu*	Remote villages
Sarawak	*Dalat*	Remote villages
Sarawak	*Daro*	Remote villages
Sarawak	*Julau*	Remote villages
Sarawak	*Kanowit*	Remote villages
Sarawak	*Kapit*	Remote villages
Sarawak	*Lawas*	Remote villages
Sarawak	*Limbang*	Remote villages
Sarawak	*Lubok Antu*	Remote villages
Sarawak	*Lundu*	Remote villages
Sarawak	*Marudi*	Remote villages
Sarawak	*Meradong*	Remote villages
Sarawak	*Mukah*	Remote villages
Sarawak	*Pakan*	Remote villages
Sarawak	*Saratok*	Remote villages
Sarawak	*Selangau*	Remote villages
Sarawak	*Serian*	Remote villages
Sarawak	*Simunjan*	Remote villages
Sarawak	*Song*	Remote villages
Sarawak	*Sri Aman*	Remote villages
Sarawak	*Tatau*	Remote villages
Terengganu	*Hulu Terengganu*	Remote villages
Terengganu	*Setiu*	Remote villages

A2: Migration matrix of population by origin and destination

Origin (1975)		Destination (1980)							Total out-migrant
		1			2	3	4	5	
		1(a)(b)	1(a)	1(b)					
	1(a)(b)	0	0	0	31,207	42,231	59,727	12,728	**145,893**
1	1(a)	0	0	116,760	19,804	25,742	28,314	5,677	**196,297**
	1(b)	0	67,070	0	11,403	16,489	31,413	7,051	**133,426**
2		55,511	30,398	25,113	0	40,534	89,852	19,426	**205,323**
3		129,205	70,247	58,958	71,777	0	168,304	95,378	**464,664**
4		213,286	105,913	107,373	171,998	184,293	0	121,714	**691,291**
5		19,901	10,913	8,988	25,282	40,370	82,888	0	**168,441**
6		32,888	17,332	15,556	17,701	16,457	22,248	19,438	**108,732**
Total in-migrant		**450,791**	**301,873**	**332,748**	**317,965**	**323,885**	**423,019**	**268,684**	

Origin (1986)		Destination (1991)							Total out-migrant
		1			2	3	4	5	
		1(a)(b)	1(a)	1(b)					
	1(a)(b)	0	0	0	39,534	45,073	62,887	10,821	**158,315**
1	1(a)	0	0	109,751	22,710	25,692	31,560	5,603	**195,316**
	1(b)	0	53,400	0	16,824	19,381	31,327	5,218	**126,150**
2		35,233	12,355	22,878	0	29,783	61,295	14,656	**140,967**
3		67,336	23,833	43,503	56,729	0	112,126	47,359	**283,550**
4		119,842	36,654	83,188	129,971	142,534	0	122,706	**515,053**
5		15,768	5,031	10,737	23,394	54,596	88,124	0	**181,882**
6		59,483	22,382	37,101	28,757	25,803	32,401	11,952	**158,396**
Total in-migrant		**297,662**	**153,655**	**307,158**	**278,385**	**297,789**	**356,833**	**207,494**	

Origin (1995)		Destination (2000)							Total out-migrant
		1			2	3	4	5	
		1(a)(b)	1(a)	1(b)					
	1(a)(b)	0	0	0	32,743	46,956	47,900	7,225	**134,824**
1	1(a)	0	0	131,112	19,016	23,995	22,766	3,645	**200,534**
	1(b)	0	27,787	0	13,727	22,961	25,134	3,580	**93,189**
2		33,438	8,440	24,998	0	22,329	45,419	8,741	**109,927**
3		63,528	16,735	46,793	42,739	0	76,796	23,362	**206,425**
4		95,412	22,870	72,542	92,831	103,826	0	52,908	**344,977**
5		19,770	4,762	15,008	19,446	43,851	52,907	0	**135,974**
6		104,110	25,693	78,417	68,065	32,445	31,821	11,111	**247,552**
Total in-migrant		**316,258**	**106,287**	**368,870**	**255,824**	**249,407**	**254,843**	**103,347**	

Origin (2005)		Destination (2010)							Total out-migrant
		1			2	3	4	5	
		1(a)(b)	1(a)	1(b)					
	1(a)(b)	0	0	0	24,900	36,942	49,661	11,269	**122,772**
1	1(a)	0	0	103,735	12,629	16,672	17,893	4,399	**155,328**
	1(b)	0	24,320	0	12,271	20,270	31,768	6,870	**95,499**
2		35,002	5,154	29,848	0	27,926	54,108	12,977	**130,013**
3		63,689	9,474	54,215	37,040	0	97,868	30,792	**229,389**
4		79,439	10,781	68,658	75,791	107,938	0	55,853	**319,021**
5		14,465	1,684	12,781	15,586	38,931	52,049	0	**121,031**
6		159,965	18,522	141,443	59,410	40,300	42,281	19,554	**321,510**
Total in-migrant		**352,560**	**69,935**	**410,680**	**212,727**	**252,037**	**295,967**	**130,445**	

Note:

1. *1 and 1(a)(b) (Capital metropolitan), 1(a) (Capital core), 1(b) (Capital suburb), 2 (Regional metropolitan), 3 (Intermediate cities), 4 (Small towns/ villages, and 5 (Remote villages).*
2. *Migrant who migrated within settlement is excluded*

A3: Origin-destination flow map, 1975-2010

Origin-destination flow map, 1975-1980

Origin-destination flow map, 1986-1991

Origin-destination flow map, 1995-2000

Origin-destination flow map, 2005-2010