Impact of Dubai Metro on Property Values

Sara Ishaq Mohammad

Supervised by: Prof. Daniel Graham and Dr Patricia Melo

Imperial College London

Department of Civil and Environmental Engineering

Doctorate of Philosophy

2014

ABSTRACT

Despite the large number of case studies estimating the impact of railways on property values, there is as yet no research conducted specifically for the Dubai Metro. This study applies the existing empirical methods, corrected for sources of bias, to test for the effect of the Dubai Metro on the values of residential and retail properties. The results will also be of interest to neighbouring cities developing their first railways.

The existing empirical work reveals a large variation in estimates for the effect of railways on property values. This thesis provides a comprehensive meta-analysis examining the sources of variation and relating these findings to the estimates for the Dubai Metro. As an addition to the existing literature, this study examines the effect on sale and rent values using repeated cross-sectional and pseudo panel data, and makes a case for the preferred data structure.

Besides applying the known measures of accessibility to a metro, this research represents the first attempt to test for the effect of the change in the generalized cost of travel (GC) due to the operations of the metro**.** The results from the preferred models indicate negative, insignificant and positive impacts of the metro on the sale value of dwellings located at different distances. The metro also enhances the rent value of dwellings and the sale value of retail properties. The study also finds an increase in the value of dwellings due to a decrease in the GC. The results suggest that while a reduction in the GC of public transport boosts the sale value of retail properties, values are higher in areas with higher private and public transport trip rates. The positive effect of the metro implies that a value capture mechanism can be explored, provided the related policy implications are understood.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

DECLARATION OF ORIGINALITY

I, Sara Ishaq Mohammad, declare that this thesis is my own work. I have exerted the best possible effort to conduct this study and presented the outcomes based on the available information and to the best of my knowledge. In addition, the work of others is referenced properly at the right positions.

COPYRIGHT DECLARATION

ACKNOWLEDGEMENT

I would like to express my appreciation to all those who have supported me in producing this piece of work. First and foremost, I express my gratitude to my supervisors, Professor Daniel Graham and Dr Patricia Melo, for their continuous support, advice and valuable feedback. The experience I have gained throughout my study has been considerable, both in respect to my research skills and my interpersonal skills.

In addition, special thanks are due to the Real Estate Regulatory Authority in Dubai and REIDIN property consultancy for collecting and providing property data. Also I thank Dubai Statistics Center for providing property data, as well as other factors such as population, employment and shopping densities.

I would also like to sincerely thank the Roads and Transport Authority in Dubai for providing transport data and all the statistics related to transport in Dubai that are considered in this study. In particular, I thank H.E. RTA's Chairman for providing support, the CEO of Strategy and Corporate Governance, the Director of Strategic Planning Department, and his colleagues in that Department as well as the Rail Agency for providing transport data and support throughout the project.

I also owe a debt of gratitude to the Emirates Foundation for funding my studies. Also, the continuous support of my family throughout my studies has been tremendous. I would like to offer them my heartfelt thanks for all their help and positive energy. Last but not least, I would like to acknowledge my friends, who have been there when needed and who have provided their fullest possible support.

Chapter 1. BACKGROUND TO THE RESEARCH

1.1 Introduction

The provision of appropriate transport infrastructure is one of the main drivers for the success of cities. The increased accessibility to neighbourhoods that transport provides brings wide economic benefits. Accessibility does not only improve commuting time, but can also lead to enhancements in trade, increases in agglomeration, improvements in land use distribution and, in many cases, increases in land and property values[1] (e.g. Banister, 2007; Debrezion et al., 2007; Dickens, 1992; Gargiulo and De Ciutis, 2010; Gibbons et al., 2010; Medda, 2012).

The benefits of accessibility may differ according to the type of transport system. While access by road is almost always ranked the highest in terms of accessibility, the gains afforded by public transport services are generally less, as well as differing by type (e.g. bus, tram, metro or marine transport). In particular, the literature suggests a large variation in the effect of railways on land and property values (Debrezion et al., 2007; Mohammad et al., 2013). While the majority of researchers report a positive effect (e.g. Agostini and Palmucci, 2008; Gibbons and Machin, 2005; Laakso, 1992; Martinez and Viegas, 2009; Pan and Zhang, 2008), some suggest that properties near stations experience a depression in value due to the increase in noise levels, pollution and crime rates (Du and Mulley, 2006; e.g. Hui and Ho, 2004), while a small number of studies imply no significant impact (e.g. Clower and Weinstein, 2002).

This chapter is structured as follows. Section 1.2 articulates the motivation for this research. Section 1.3 discusses the scope of the study, while the research aim and objectives are listed in section 1.4. In section 1.5 we discuss the methodological challenges in estimating the effect of railways on land and property values, and the possible approaches to address these issues. The contributions of this research are presented in section 1.6. Finally, we provide the outline of the thesis in section 1.7.

---

[1] The term land refers to an empty plot of land, whereas property refers to land that has been built on.

1.2 Background

Examining the effect on the value of properties of accessibility to transport systems in general, and to railways in particular, does not only provide an assessment of the impact of the transport system, but is also worthwhile for decision makers involved in the recommendation or implementation of transport solutions (e.g. Billings, 2011; Bowes and Ihlanfeldt, 2001; Du and Mulley, 2006). For example, understanding the economic impact of an improved or a new railway versus that of a bus system can affect the choice of the most suitable transport solution for the development plans of a city. In addition, some researchers support capturing at least part of the accessibility benefits on property values to fund these projects and different mechanisms to achieve this have been proposed in the literature (Medda, 2012; Smith et al., 2010).

Enoch (2002) states that the positive effect of accessibility to rail stations was first recognised at the time of the construction of the London Underground in the 19^{th} century, when land in the vicinity of railway stations was sold at higher market values. Attempts to measure the extent and the magnitude of the impact only started in the 1960s, however. Since then, the majority of the research has been for cases in the United States of America and, recently, in Europe, East Asia and a few other parts of the world. With the opening of the Dubai Metro in September 2009, Dubai introduced the first railway system in the Middle East. The metro consists of the Red and Green Lines, which started operations in September 2009 and September 2011, respectively. The two lines serve the areas with the highest trip demands in Dubai. This is the first study to shed light on the impact of the metro on the value of residential and retail properties in Dubai.

The motivations for this study are diverse. First, estimates of the effects of the Dubai Metro on property values represent the first objective assessment of the metro. Second, the results provide a basis for decision makers in Dubai to develop property value capture policies to assist the funding of future metro lines. Together, these two motivations also support the decision process in terms of whether to expand the railway system in Dubai. Third, since many cities in the Middle East with similar characteristics to Dubai are also planning their first metro system, this study provides the most relevant benchmark for these cities in terms of the expected effect of the metro on property values.

1.3 Scope of the research

The research examines the effect of the operations of the metro Red Line on the value of residential and retail properties in Dubai. Out of nine obtained land and property datasets, four which refer to sale transactions and sale listings[2] of dwellings, rent listings of dwellings and sale transactions of retail properties, are suitable and considered for analysis. While transaction records are available for the years 2007 to 2011, listing records consist of property data from 2009 to 2011.

This study does not only consider the impact of accessibility to a metro station on the value of properties, but also tests for the effect of changes in the generalized cost (GC) of travel due to the operations of the metro. Generalized cost of travel is tested using both public transport systems (metro, public buses and marine transport) and private transport systems (cars, vans and taxis). The implications of the metro for property values is discussed in terms of value capture policies, but the study does not itself develop such funding mechanisms for Dubai.

1.4 Research aim and objectives

This research aims to examine the impact of accessibility to metro stations using different measures and different empirical methodologies to establish causality. The objectives of the research are to:

1. Summarize critically and objectively the literature on the impact of railways on land and property values.
2. Enrich the existing property data by adding missing attributes that might affect property values, such as building attributes and neighbourhood characteristics.

[2] Listing values are the prices that a property owner asks for. The words listings and asking values are used interchangeably in this study.

3. Estimate the impact of the Dubai Metro on the sale and rental values of residential properties (transactions and listings) and retail properties (transactions). This is achieved by using the following:
   3.1 Two data structures (repeated cross-sectional data and pseudo panel data[3])
   3.2 Different empirical strategies considering causal relationships.
   3.3 Three measures of accessibility benefit to the metro (a binary measure, accessibility to a number of stations and the change in the GC of travel due to the metro).

1.5 Methodological issues and challenges

The main methodological issues in estimating the effect of railways on land or property values are reverse causality, confounding or omitted variable bias (OVB), unobserved heterogeneity, and measurement error. Here we briefly introduce the potential causes of bias and the measures adopted to address them.

Reverse causality occurs if the dependent variable affects one of the independent variables, leading to biased estimates. In the case of this study, the choice of the railway alignment or the location of stations may be affected by the value of properties. For example, values can be higher in areas with higher access to a railway for reasons other than accessibility to a station (e.g. due to lower levels of noise and pollution). In models that do not control for causality, therefore, the coefficient estimate measuring the effect on property values of access to the rail network is likely to be over-estimated.

Omitted variable bias, or confounding, occurs if one or a number of significant variables are intentionally or unintentionally left out of the regression model. This leads to a correlation between an independent variable and the error term, hence biasing estimates. The potential for confounding largely depends on the availability of data and the judgement of the researcher with

[3] Repeated cross-sectional data contains at least one sale and one resale of the same clusters of properties over a period of time. Pseudo panel data is semi-panel data constructed by grouping the same observations over a period of time. More details on how the pseudo panel data is constructed in this study are provided in chapter 5.

regard to providing the most extensive available set of variables in the models. For example, there may be fundamental differences between the characteristics of neighbourhoods served by a railway compared to unserved neighbourhoods, and these differences need to be controlled for.

Estimates of the impact of railways on land or property values can also be biased in the presence of uncontrolled unobserved heterogeneity across land or properties. For example, it is possible that the quality of properties in the vicinity of railways may be different than the quality of properties located further away. Since quality is not generally measured, failure to control for unobserved heterogeneity increases the error in the model and leads to potentially inconsistent and biased estimates. Finally, measurement error occurs if the values of the independent variables are falsely measured. This biases the findings since it leads to a correlation between the independent variable and the error term.

Different methodological challenges are associated with different empirical methods, and the data structure used to estimate the effect of railways also has an influence on the potential sources of bias. In this study, we aim to reduce bias by adopting a variety of empirical methods, including difference-in-differences (DID) and hedonic pricing (HP) corrected for bias, as well as using two different data structures (repeated cross-sectional and pseudo panel data). These regression models relate land or property value to various contextual factors including proximity to a railway. The models are explained further in the next chapter.

Reverse causality can be controlled through a combination of data structure and empirical method. For the data structure, one needs to ensure that either the same properties, or the same groups or clusters of properties, are observed before and after the construction of the railway. For the empirical method, as an improvement to the widely used HP models, this study adopts the DID method, since this addresses reverse causality.

Additionally, controlling for the largest possible list of contextual factors that might be expected to affect the value of land and property allows one to control for OVB in the HP models. This study checks for confounding variables and includes them in the empirical analyses. In addition, by tracking properties over time, the DID method reduces omitted variable bias. Bias from unobserved heterogeneity can be reduced by tracking for the same land or property or the same group of land or property over time.

Finally, measurement error depends largely on the data structure. In this study, measurement error is not an issue in the repeated cross-sectional data because a reasonable level of accuracy in the observed independent variables can be assumed in this dataset. The pseudo panel data, however, is more prone to containing measurement error, especially when the number of observations in a constructed property group is not sufficiently large.

1.6 Research contributions

This study contributes to the existing literature in five main areas. Firstly, very few studies have attempted to analyse the large variation in results across case studies of the effect of railways on land and property values. Such attempts at comparison as there are have either relied on a subjective comparison of findings (RICS Policy Unit, 2002; Ryan, 1999) or considered case studies from one geographical region (Debrezion et al., 2007). This research, however, conducts a much more comprehensive meta-analysis in order to examine the widest possible range of findings. We provide more discussion of the contributions of this meta-analysis to the literature in chapter 3. In addition, the results of the meta-analysis are used to extract factors that affect property values and also to discuss the findings from this study.

Secondly, this study provides the first objective estimates of the impact of the Dubai Metro on residential and retail property values by applying existing methodologies (HP and DID) in a new context. As discussed in section 1.2, the findings are not only of interest to Dubai but also to similar cities building their first railway and to policymakers seeking means of recovering at least part of the anticipated value increase due to the accessibility offered by the metro.

Thirdly, previous research has generally used one type of data structure to estimate the effect of railways on land and property values. This study, however, uses two data structures (repeated cross-sectional data and pseudo panel data) and argues for the most suitable data arrangement for cases such as these.

Fourthly, since only a limited database of property transactions and asking values was available in Dubai, the datasets were heavily enriched in this study to combine property and building data, transport data, neighbourhood characteristics and accessibility to amenities. The majority of the variables were collected manually for the purpose of the research and these rich datasets can be used in many other studies in Dubai.

In addition, there is a gap in the literature regarding the estimation of the effect of the change in overall travel costs due to the operations of a railway (e.g. Bae et al., 2003; Dewees, 1976; Ryan, 1999; Vichiensan and Miyamoto, 2010). For example, Ryan (1999) states that the calculation of the accessibility benefit should be based on the reduction in travel costs. Some studies that have examined the impact of railways on other economic factors have also concluded that there is a need for a more robust measure of accessibility. Gibbons et al. (2010), for example, state that "better data on transport cost changes induced by transport projects is also highly desirable". One way to measure such transport cost changes is to estimate the impact of the change in the generalized cost of travel due to the introduction of the metro. This approach, which is adopted in this study, is, to the best of our knowledge, original, and marks the fifth contribution of this research. This method can also be used to measure the impact of any type of transport system.

1.7 Thesis outline

The remainder of the thesis is structured as follows. Chapter 2 reviews the previous empirical work related to the effect of railways on land and property values. The chapter starts by explaining the theory of interaction between transport accessibility and land and property values. Next, we summarize the empirical evidence from previous studies, including the results, sources of bias, methods and data structures. This is followed by the presentation of the factors that affect the value of land and property. Finally, one of the main motivations for estimating the effect of railways is explored through an introduction to the fiscal mechanisms used to fund railway schemes.

While the second chapter offers a subjective review of previous empirical work, a more objective review is conducted in chapter 3. After defining the scope of the meta-analysis, the model design is described, which is then followed by the results. The importance of the meta-analysis to this study is also outlined.

Before proceeding with the data and the analysis of the impact of the Dubai Metro on property values, the emirate of Dubai and the Dubai Metro are put into context and introduced in chapter 4. The chapter provides, firstly, background information on Dubai's growth and urban structure, secondly, a description of the development of the transport supply and the changes in travel demands, and finally, an introduction to the Dubai Metro, the choice of metro routes and the characteristics of the communities it serves.

Following this, chapter 5 covers the study design rules by discussing the data structure as well as the logic behind the choice of the empirical methods for this study. To begin with, the properties affected by the metro (treated) and the properties unaffected (control) are defined. This is followed by the definition of the treatment year. Next, we explain the construction of the pseudo panel data. Finally, the empirical methods used in the study are discussed and justified.

Chapter 6 explains the transport data that was obtained and describes and analyses the property data. The transport data consists of the variables used to measure the accessibility benefit offered by the metro. The chapter also discusses the consistency of the original property datasets, the refinement of the observations and the selection of the final datasets, as well as providing a description of the available and added attributes to the original datasets. Since this study generates results using two data structures, we also provide descriptive statistics of the property data related to each structure.

The design of the models and the estimates using the two data structures are covered in chapters 7 to 9. Chapter 7 starts by explaining the basis of the DID empirical method that was developed for estimating the effect of proximity to Dubai Metro on the value of properties. The empirical results for each of the four property datasets are presented.

Chapter 8, meanwhile, develops the HP models that are generally used in the literature and presents the consequent results for the effect of the metro on the sale and rental values of dwellings and the sale value of retail properties.

The third empirical chapter (chapter 9) examines the effect of the change in the generalized cost of travel due to the metro. The gap in the literature in estimating the overall effect of a railway on the travel times and costs is explained. The model design is then developed to test for the impact of the changes in the generalized cost of travel using public and private transport systems. The results are then presented and discussed separately for each dataset.

The thesis concludes with chapter 10, which outlines the key findings of the study and discusses the differences between the models and data structures. In addition, this chapter discusses the policy implications of the results, the contributions made by the study, its limitations and proposals for future research, both in Dubai and in the transport field in general.

Chapter 2. LITERATURE REVIEW

This chapter provides a review of the literature related to the effect of railways on land and property values. Section 2.1 explains the interaction between transport and land and property values, focusing on the most common theory in this regard. This is followed, in section 2.2, by a summary and a discussion of the empirical evidence regarding the change in land and property values due to railways. This section provides background on the previous research conducted in relation to the Dubai Metro, sets out the variation in land and property values across case studies, discusses sources of bias in estimating the effect of railways on land and property values and addresses the empirical methods and data structures used in the literature to reduce bias. Next, in section 2.3, the factors that affect land and property values are listed and described. Section 2.4 reviews briefly the fiscal mechanisms used to capture the enhancement in land and property values due to transport, and the overall conclusions drawn from the literature, and the lessons learnt are discussed in section 2.5.

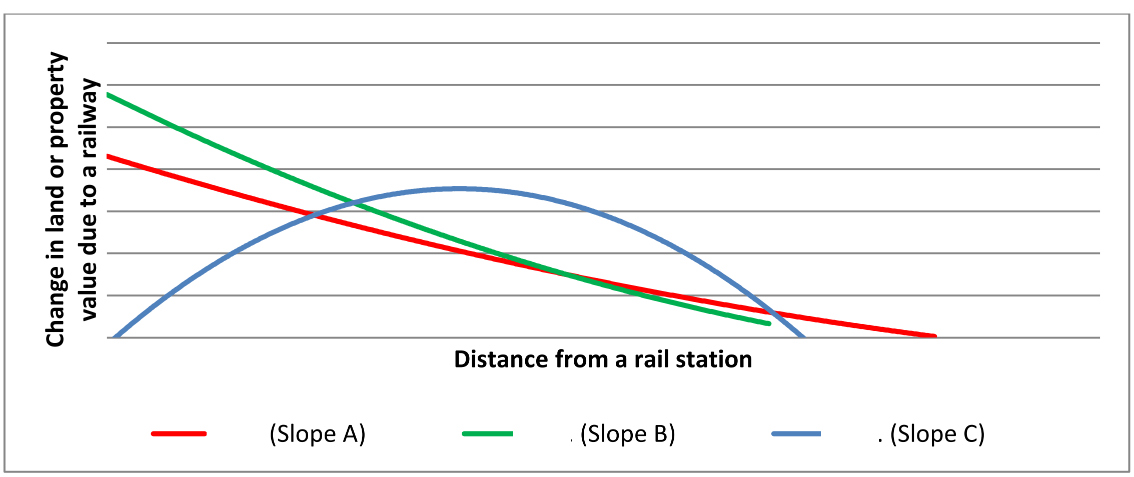
2.1 Theories of interaction between transport and land and property values

Location theories explain the variation in economic activities due to their geographical location. While such theories consider a range of factors that can impact location, optimizing transport costs is generally the factor with the largest impact weight. Although there are other uses for location theories, the two main areas in which location theory is used are land use location (introduced by Von Thunen) and manufacturing industrial location (introduced by Weber) (Fujita, 1989; Isard and Peck, 1954; Krugman, 1993; Predöhl, 1928). Debrezion et al. (2007) argue that the theories that explain the relationship between transport and the location of land or property tend to take one of two directions in the literature; the first tests for the optimal location of land or property considering a number of factors, including transport, while the second examines the relationship between location and the value of land or property. Since the purpose of this study is to estimate the effect of the Dubai Metro on property values, this discussion focuses on explaining the second type.

The most common location theory to explain the variation in land and property values is called the bid-rent theory (Debrezion et al., 2007; Duncan, 2008; Martinez and Viegas, 2009; O'Sullivan, 2003; Ryan, 1999; Zhang, 2009; Fujita, 1989). Bid rent theory can be traced back to the nineteenth century scholar, Johann Heinrich Von Thunen, who sought to explain the accessibility of farmlands to business areas. Von Thunen's theory suggested that the ability of a transport service to reduce the costs of transporting agricultural products to business areas plays a major role in the valuation of farmlands. Alonso (1964) and Muth (1969) developed this theory further to a bid rent theory that relates the bid value of land and properties to their accessibility to the Central Business District (CBD). This assumes a mono-centric city in which the CBD is the most highly valued area and where accessibility to the CBD plays a major role in the value of land and properties. Similarly, the value of land and property can be affected by proximity to shopping areas and accessibility to a transport system (e.g. a rail station or a highway).

The slope and the magnitude of the bid rent curve vary according to one, or a combination of, factors such as property type, neighbourhood income, attractiveness of the transport system and the location of land and property within the city (e.g. Baum-Snow and Kahn, 2000; Debrezion et al., 2007; Fujita, 1989; McCann, 2001; Mohammad et al., 2013). Figure 2-1 is a symbolic demonstration of the concept of this theory in the context of this study. The slopes show the most common examples of the effects of accessibility with increasing distance to a railway station, and these are explained below in relation to the factors that can affect the impact rate and radius.

Figure 2-1: A demonstration example of the bid-rent theory (source: self-produced graph)



The literature shows that, in general, the effect of accessibility to a railway on the value of residential properties extends further than for commercial properties[4], and the slope of the curve is also shallower (O'Sullivan, 2003). This suggests that residents are willing to commute further to reach a railway station compared to those commuting to a commercial area. Nevertheless, the theory also suggests that the rate of the impact of accessibility to a railway is higher on the value of commercial land or properties than it is for residential land or properties. In the above figure, the slope for the residential and commercial properties can be marked as A and B, respectively.

The curve may also vary according to the income level and the accessibility preferences of commuters. The theory suggests that in urban areas, low income people tend to be cautious about their travel expenditure. As the cost of using public transport services are generally lower than that for private transport, all else being equal, the value of accessibility to railway stations is therefore higher in low income neighbourhoods, and the slope of the curve is steeper, compared to that for properties located in high income neighbourhoods (O'Sullivan, 2003). For example, Du and Mulley (2006) found that for properties located at the same distance from a station, while the value of dwellings in one neighbourhood reduced, the value of properties in another neighbourhood increased. McCann (2001), however, argues that the situation reverses in cases where individuals place a higher value on accessibility to railway stations, meaning that the value of properties near a station is bid higher and therefore higher income people tend to live nearer stations.

The demographic characteristics of the population in a given neighbourhood can also influence the extent and magnitude of the effect; for example if an area is occupied by younger people and they happen to prefer railway services over other transport services, then the accessibility to railway stations tends to uplift property values more (McCann, 2001). In addition to the above, the bid-rent curve can vary according to the impact of negative externalities (e.g. crime levels) and the importance of the environmental conditions of an area. In general, the higher the income-

[4] Commercial properties refer to retail and offices

level of a neighbourhood, the higher the importance of the environmental conditions (i.e. low levels of noise and pollution).

In some cases, the bid rent curve can be concave (slope C); indicating that land and property values reduce at closer distances to railway stations due to the increase in negative externalities, so that the positive impact reaches a peak at a larger distance from a station and fades away later (McCann, 2001). Finally, the effect may also vary with time or the maturity level of the railway. For example, the short term impact of proximity to a rail station may be concave (slope C), but as the system becomes more popular in the long term, the effect can move to a reducing slope (slopes A or B).

Recent advances have not only considered the theory that explains the accessibility offered by a transport service, however, but also the potential impact of all other contextual factors on land or property values. These approaches have gained popularity after the development of a multivariate regression model of hedonic pricing in the work of Rosen (1974). The next section describes the evidence from previous empirical work, recent advances in regression methods and data structures.

2.2 Empirical evidence from previous studies

The theory of interaction between transport and land and property values suggests that the nature of the impact (i.e. the rate and the radius) differs between case studies. This section summarizes the literature on the effect of railways around the world on the value of land and property. Previous research on the impact of the Dubai Metro on the value of land and property is presented first. The reported empirical results in terms of the extent and the magnitude of the impact of railways, and their relation to a number of contextual design factors is then discussed, followed by an assessment of the sources of bias in estimating the effect of railways on land and property values. This is followed by a review of and a comparison between the empirical methods. Finally, the type of data structure and its effect on the reported estimates is examined.

2.2.1 Previous research on the impact of the Dubai Metro on land and property values

Minimal research has been done to identify the impact of the newly opened Dubai Metro on land and property values and, to the best of our knowledge, no study has used a statistical method to estimate the effect. This sub-section discusses the previous studies in this regard.

The first study was a qualitative review report, that was prepared before the metro started operations, by the consultant DTZ on the perceived impact of the metro on property values (Edwards and Cooper, 2009). The report discusses various factors that may affect the sale and rental values of properties near metro stations; these are the property location, land use type and the characteristics of the metro. The real estate consultant only subjectively predicted that the metro may enhance the value of commercial and residential properties located within 400 m and 1 km of a station, respectively, by a range of 10% to 20%. The study also mentions that the value of properties in the more employment dense areas may experience a considerable uplift.

The second analysis was part of a study commissioned by the Roads and Transport Authority (RTA), which aimed to establish a methodology to collect at least part of the accessibility benefit to public transport services from the possible increase in land and property values so as to fund future rail projects (MVA, 2011). One aim of the study was to identify a tool to calculate the impact of the Dubai Metro on the value of land and property around stations. The study mentioned various regression methods to conduct the analysis, but since these methods are data and resource hungry they were therefore considered to be beyond the scope of work of the project, and were not further considered. The proposed methodology was to relate land and property value enhancement in close proximity to metro stations to other benefits gained, such as reduced parking plots and savings in the demand for private transport. These benefits, however, are not necessarily directly related to the effect of the metro on land and property values.

As part of the same study, RTA conducted a survey with a number of property advisers and real estate agents in Dubai in order to capture their views on the perceived effect of the metro on property values before services started. The survey report mentioned that the interviewers anticipated that the metro could increase residential property values by 10-30%, but that the value of dwellings located very close to station may decrease by up to 10%, or not be affected at all, due to the increase in the level of noise and reduced environmental conditions around stations. For commercial properties, a 10-40% increase was anticipated. Finally, neighbourhoods suffering congestion were expected to benefit the most.

2.2.2 Empirical results

A large number of studies have analysed the impact of a new or improved transport scheme on land and property values. There are a larger number of studies estimating the effect of railways on residential property values compared to studies on other property types and land uses. This is due to the larger number of available transactions for the residential market (Diaz, 1999; Duncan, 2008). A summary of the main outcomes and design context of these studies are provided in

Table 2-1. We have also estimated the annualized effect of the railway in each case study.

While the majority of estimates suggest a positive impact (e.g. Agostini and Palmucci, 2008; Laakso, 1992; Pan and Zhang, 2008; Voith, 1991), it can be observed that the effect ranges greatly from positive to negative. In some cases, the effect varies across the city as well. For example, the study by Du and Mulley (2006) on the effect of the Tyne and Wear Metro in England suggests that the effect of proximity to the light rail stations varies across the city. The authors use records for one month and find that while the metro boosts values of residential properties in some locations, it reduces the value of properties located within a catchment area of 200 m to 500 m in a non-CBD area by between 5% and 40%. The authors explain that an understanding of the difference in the neighbourhood characteristics in these areas can help explaining the variation in results. In addition, they argue that positive enhancements can occur across the city after a longer time of operations than that used in the study.

A few other studies also indicate that the land or property in close proximity to a rail station experiences a fall in value due to negative externalities such as increases in noise, pollution and crime levels (e.g. Diaz, 1999; Hui and Ho, 2004; Brinckerhoff, 2001). Bollinger et al. (1998), for example, found a negative effect over 6 years on the rental value of offices located within quarter of a mile of Atlanta Rail stations (USA). Although the authors relate this only to the reduced safety around the stations, the competiveness of railway services in relation to other modes may have affected the results, since the same study also suggests that the proximity to the highway raises the value of offices. This in turn can indicate that commuters in this city appreciate access by private transport mode more than by public transport services. The attitudes towards, and willingness to use, public transport are vastly different across cities and this can impact on the estimates.

A small number of case studies suggest no significant impact of railways on land and property values (Gatzlaff and Smith, 1993). The insignificant effect of railways can be explained by two main reasons; either the perceived benefit of accessibility to a railway station (the saving in travel times and costs) is not achieved (Debrezion et al., 2007; Baum-Snow and Kahn, 2000; Dubé et al., 2013) or the positive and negative effects of the railway balance each other out (McCann, 2001). In the absence of clear benefits to users of and commuters to an area, the price that consumers are willing to pay for a land and property may not be affected by proximity to a new or an improved railway.

Some studies have stated that the ability of an improved or new railway to increase property values lies in the ability of the new system to reduce the overall travel times and costs, despite the fact that other negative externalities may have a dampening effect (see Bowes and Ihlanfeldt, 2001; Duncan, 2008; Ryan, 1999). In other words, measuring accessibility only by distance, or time, or as a binary value of proximity to a railway station, may not truly represent the full accessibility benefit. There is a need, therefore, for the literature to measure the effect of the saving in travel times and costs due to a transport service on property values (Ahlfeldt, 2013; Bae et al., 2003; Dubé et al., 2013; Gibbons et al., 2010; Ryan, 1999).

The main differences across case studies are summarized below in terms of the type of property, type of railway, level of maturity of rail service, the geographical region of the case study and the impact radius. Results across studies suggest that the effect of railways on the value of commercial properties is generally higher than the effect on the value of residential properties (see also Debrezion et al., 2007; Mohammad et al., 2013). In addition, railways are estimated to have a greater impact on land values compared to property values. Clower and Weinstein (2002) also suggest that the positive effect on values is greater in more congested areas.

It can also be noted that the impact varies according to the type of railway; while the impact of proximity to a light railway or a metro station ranges between negative and positive, the effect for commuter rail is positive in the majority of cases. In addition, the effect may be dependent on the level of maturity of the service and its attractiveness and competiveness compared to other transport services. Two studies of the same railway conducted at different maturity levels provide evidence for this argument. In a study on the effect of the Metropolitan Atlanta Rapid Transit rail on property values, Nelson and McCleskey (1989) report no significant impact after a few years of operations, whereas the study conducted by Bowes and Ihlanfeldt (2001) at a later date suggests positive and negative effects at different locations. Similarly, the effect of the BART system in San Francisco on property values capitalized after a period of time (Clower and Weinstein, 2002).

In addition, estimates can vary according to the geographical location of the railway. The effect of proximity to railway stations in most European and East Asian cities is higher compared to cities in North America. This may be due to the higher trip share by public transport services in Europe and East Asia compared to the more car-oriented cities in North America. The RICS Policy Unit (2002) states that the effect of public transport on property values reaches its positive peak in public transport oriented cities and in cities with restricted car access or low car share.

The area over which the influence of railways on the value of land or property extends varies according to the land use type. Studies suggest that the impact radius is larger for residential properties than for commercial properties (Cervero and Duncan, 2002; Debrezion et al., 2007). The RICS Policy Unit (2002) suggest that the effect can go beyond 2 km in residential areas, compared to 1 km in commercial areas. Nonetheless, the impact is not linear with distance from a station, but rather decays like the bid rent theory (e.g. Al-Mosaind et al., 1993).

Table 2-1: Example summary of property value changes in close proximity to rail stations (source: self-produced table)

Author(s)	Type	Rail system	Location	Estimation method*	Estimated effect	Goodness of fit (R^2)	Data time span**	Time horizon compared to start time of rail operations	Average annualized effect
	Properties								
Laakso (1992)	Residential	Metro	Helsinki, Finland	Hedonic pricing	3.5% to 6%	0.94	3 years	2 years before, 6 years after	0.9% to 1.5%
Al-Mosaind et al. (1993)	Residential	Light rail	Portland, USA	Hedonic pricing	10.6%	0.60	1 year	2 years after	10.6%
Benjamin and Sirmans (1996)	Residential	Metro	Washington, DC, USA	Hedonic pricing	2.5% due to 10th a mile distance reduction	0.74	1 month	16 years after	0.16% due to 10th a mile distance reduction
Chen et al. (1997)	Residential	Light rail	Portland, USA	Hedonic pricing	10.5%	0.63	4 years	2 to 6 years after	2.6%
Bollinger et al. (1998)	Office	Light rail	Atlanta, USA	Hedonic pricing	-7%	0.63	6 years	6 years after	-1.2%
Weinstein and Clower (1999)	Residential	DART light rail	Dallas, USA	Average comparison of value change	-5.2%	Not reported	5 years	2 years before, 3 years after	-2.1%
	Retail				4.6%				1.8%
	Office				22.7%				9.1%
FTA (2000)	Commercial	Metro	Washington, DC, USA	Hedonic pricing	2% increase in value for every 1000 feet reduction in distance	0.11 to 0.38	2 years	3 years before, 3 years after	0.7% increase in value for every 1000 feet reduction in distance
Chesterton (2000)	Residential	Underground	London, UK	Hedonic pricing	71.1% & 42%	F statistics: 9 to 90	10 years	5 years before, 5 years after	14.2% and 8.4%
Bowes and Ihlanfeldt (2001)	Residential	MARTA	Atlanta, USA	Hedonic pricing	-19% to 2.4%	0.48	4 years	8 to 12 years after	-1.9% to 0.24%

Weinberger (2001)	Office	Light rail	Santa Clara County, USA	Hedonic pricing	7% to 10%	0.73 to 0.79	16 years	3 years before, 13 years after	0.8% to 1.3%
Clower and Weinstein (2002)	Residential	DART light rail	Dallas, USA	Average comparison of value change	7.2% & 18.2%	Not reported	5 years	2 years before, 3 years after	2.9% to 7.3%
Bae et al. (2003)	Residential	Seoul's rail	Seoul, Korea	Hedonic pricing	0.13 to 2.6%	0.96	4 not continuous years	8 years before, 4 years after	0.02% to 0.4%
Cervero (2003)	Residential	Light and commuter rail	San Diego County, USA	Hedonic pricing	-12% to 46%	0.61 to 0.74	1 year	during opening year	-12% to 46%
	Commercial				71.9% to 91%	0.83	3 years	1 year before, 2 years after	24% to 30.3%
Yankaya and Celik (2004)	Residential	Metro	Izmir, Turkey	Hedonic pricing	0.7% & 13.7%		4 months	4 years after	0.4% to 6.9%
Gibbons and Machin (2005)	Residential	Underground	London, UK	Differences in differences and hedonic pricing	Minimum of 1.5% increase in value for every 1 km distance reduction	0.75 to 0.89	5 years	3 years before, 2 year after	Minimum of 0.6% increase in value for every 1 km distance reduction
Debrezion et al. (2006)	Residential	National railway	Holland	Hedonic pricing	25%	0.82	Not provided	cannot be measured	Not provided, cannot be measured
Du and Mulley (2007)	Residential	Tyne and Wear Metro	England, UK	Geographically weighted regression	-42% to 50%	F value 5.39	1 month	22 years after	-3.8% to 4.5%
Duncan (2008)	Residential	Light rail	San Diego, USA	Hedonic pricing	5.7% & 16.6%	0.83 to 0.86	5 years	2 to 7 years after	1.3% to 3.7%
Pan and Zhang (2008)	Residential	Shanghai rail transit system	Shanghai, China	Hedonic pricing	1.1% & 3.3%	Not reported	Not reported	Not reported	Not reported and cannot be measured

Agostini and Palmucci (2008)	Residential	Santiago metro	Santiago, USA	Differences in differences and hedonic pricing	From 3.8 to 7.4%	0.70	4 years	4 years before	0.95% to 1.9%
	Land								
Weinstein and Clower (1999)	Residential	DART light rail	Dallas, USA	Average comparison of value change	7.7%	Not reported	5 years	2 years before, 3 years after	3.1%
	Retail				29.7%				11.9%
	Office				10.1%				4%
Cervero and Duncan (2002)	Commercial	Light rail	Santa Clara County, USA	Hedonic pricing	23%	0.31	15 years	8 years before, 7 years after	3.1%
		Commuter			120%	0.31			16%

* More information on estimation methods is provided in section 2.2.4

2.2.3 Sources of bias in estimating the effect of railways on land and property values

The empirical models used to test for the effect of a railway on the value of land or property can potentially suffer from bias. The four main sources of bias are reverse causality, omitted variables (OVB), unobserved heterogeneity across properties and measurement error (e.g. Billings, 2011; Dubé et al., 2013; Gibbons and Machin, 2005; McDonald and Osuji, 1995; McMillen and McDonald, 2004). These issues are discussed in more detail in this section, while the following two sections describe the empirical models and the data structures that can address these methodological challenges.

Reverse causality is the possibility that the dependent variable may have affected one of the independent variables (Abadie, 2005; Pearl, 2009). For example, the value of properties in an area may have attracted the railway in the first place; hence reverse association between property value and accessibility to the railway needs to be controlled for to avoid biased estimates.

While all regression based models find an association between the dependent variables and the independent variable, a few consider the reverse relationship. This is addressed by comparing the values for the properties affected (treated) and those unaffected (control) by the railway and for the period before and after the railway is introduced or improved. The results are unbiased given that the value changes across properties remain similar over time in the absence of the railway.

Another method that can address reverse causality is instrumental variables. This relies on the availability of a variable (an instrument) that induces the selection of the treatment only. In other words, the instrument is correlated with one of the contextual factors and affects the independent variable indirectly, only through the treatment. For example, the ability of a person to learn (i.e. the instrument) affects his expected salary (i.e. the independent variable) only through its correlation with the education level (i.e. the treatment). To use this method, the number of instruments must be at least equal to the number of contextual factors in the regression models. To the best of our knowledge, no study has been successful in using the instrumental variables method to estimate the effect of a transport system on the value of land or property. This is probably due to the lack of availability of the right instruments.

The other potential bias in estimating the effect of railways on property values is OVB, or confounding. This occurs if one intentionally or unintentionally omits variables that significantly affect the value of land or property. This can lead to fundamental differences in the values of land or property exposed to the railway compared to those not exposed (i.e. due to reasons other than access to a rail). For example, a simple relationship between the value of land or property (y_{it}) and an accessibility measure to a railway (X_{it}) at time t can be expressed as shown below:

$$y_{it} = \alpha + \beta_i X_{it} + \varepsilon_{it} \tag{1}$$

For the above equation to produce unbiased and consistent estimates, the error term (ε_{it}) is assumed to be independent of the accessibility measure to the railway (X_{it}), hence the error has the same mean across all property observations. Nevertheless, in the presence of confounding variables (C_{it}) that are not included in the model, the error term will be correlated with the dependent and the independent variables, leading to bias.

Recent advances in least-square regression models allow one to control for confounding by adding these variables in the model (Abadie, 2005; Frank, 2000). In the context of this research, the variables that previous researchers have identified as possible confounders are accessibility to the CBD or to employment centres (e.g. Dewees, 1976; Du and Mulley, 2007; Dubé et al., 2013; Gibbons and Machin, 2005; Kim and Zhang, 2005; Martinez and Viegas, 2009; Wu, 2012) as well as to shopping centres (e.g. McDonald and Osuji, 1995; Vichiensan and Miyamoto, 2010).

In addition to the above methodological challenges, some variables that are expected to affect land or property values may be unobserved, such as the quality of the building and the environmental conditions of a neighbourhood. Differences in these aspects between neighbourhoods served and not served by the railway lead to local specific effects related to each property, which can impact the results if not accounted for. The unobserved heterogeneity across properties leads to correlation with the error term which, as a result, biases estimates. To correct for this, observations for the same properties have to be available over a period of time so as to cancel out the effect of the time-invariant unobserved effects related to each property.

Finally, measurement error occurs where the collected or the constructed land or property observations are inaccurate. Almost all studies assume that the obtained data is measured correctly, however if the researcher chooses to re-arrange the data to account for other sources of potential bias, such as the construction of pseudo panel data that deals with unobserved heterogeneity, the refined dataset is prone to measurement error unless careful measures are considered. These are introduced further in section 2.2.5.

2.2.4 Empirical methods

There are many methods to estimate the effect of a transport system on land and property values. While the majority of researchers employ regression based methods to examine the impact, a limited number employ the rather weaker approach of a simple comparison of the average change in land or property values before and after introducing or improving the transport system. This section provides a review of the regression based empirical methods used in previous studies, focusing on the causal approaches that address the possible sources of bias.

The most widely employed method is hedonic pricing (HP) (e.g. Al-Mosaind et al., 1993; Bae et al., 2003; Bowes and Ihlanfeldt, 2001; Cervero and Duncan, 2002; Voith, 1991). HP models are used to analyse the effect of a particular treatment on a change in an economic outcome, such as a property value, employment ratio and productivity. The models have been applied extensively after the work of Rosen (1974). In cases like this study, HP models examine the impact of various factors on land or property values including land or property characteristics, location attributes and accessibility to transport. The typical form of an HP model is shown below:

$$y_{it} = \beta_0 + \sum_{m=1}^{m} \beta_m X_{mit} + \mu_{it} + \varepsilon_{it} \tag{2}$$

Where y_{it} is the value of observation i in time t, X_{mit} is the value of the covariate m that is expected to affect the dependent value y, for observation i in time t, μ_{it} controls for the unobserved factors and ε_{it} is the error term.

The advantages and limitations of the HP models are as follows. These models are well developed in the literature, provide an association between land or property values and contextual factors based on actual observations and are versatile because they allow for different combinations of variables. Some issues need to be considered, however. First, HP models lack a causal interpretation since they do not control for the possibility of reverse effect of property values on accessibility to a railway. Second, the functional form can affect the results, leading to model misspecification, although the use of semi-parametric models can solve this. Third, to reduce bias from OVB, HP models rely on high quality data and a comprehensive list of contextual factors that are expected to affect land or property values.

In addition, the researcher needs to have a prior knowledge of potential confounding variables and to check for the effect of these variables on the reported estimates (e.g. Billings, 2011; Dewees, 1976; Du and Mulley, 2007; Kim and Zhang, 2005; McDonald and Osuji, 1995; Vichiensan and Miyamoto, 2010; Wu, 2012). For example, Bowes and Ihlanfeldt (2001) use HP models to study the effect of proximity to a rail station on the value of residential properties, arguing that their results are unbiased since they control for potential confounders of neighbourhood attributes (e.g. employment density, retail employment density, median income, crime level).

While HP models provide a global estimate of the effect of a transport system on property values, a more recent methodology, known as the geographically weighted regression (GWR) method, relates the outcome variable to the location of the independent variables and calculates coefficients locally (Fotheringham et al., 2002; RICS, 2004). The GWR method was introduced by Brunsdon et al. (1998) and is used in studies with evidence of spatial variation in the effect of an economic activity on the surrounding areas (Mennis, 2006). The GWR method has the same advantages and limitations as the HP method, with the exception that it provides local estimates of the impact. A few studies apply GWR to examine the effect of a rail system on property values, differentiating their results geographically (see Du and Mulley, 2006; Saphores and Yeh, 2013; Vichiensan and Miyamoto, 2010; Zhou and Zhang, 2010).

In addition to the above, recently, some studies have tested for the relationship between accessibility to a railway and land and property values using an innovative model, referred to as the difference-in-differences (DID) model. The effect of the railway (treatment effect) is estimated by comparing the value of properties affected by the treatment (treated) to those that did not experience any effect (control), both before and after the treatment. For example, for each treated property observation "i", the value pre- and post- the treatment is defined as y^1_{i0} and y^1_{i1}, respectively. Similarly, for a control property, the value pre- and post- the treatment is defined as y^0_{i0} and y^0_{i1}, respectively. Therefore, the effect of the railway on land or property value (the treatment effect), β, is

$$\beta = (\bar{y}^1_1 - \bar{y}^1_0) - (\bar{y}^0_1 - \bar{y}^0_0) \quad (3)$$

Where $\bar{y}^1_1$ is the mean value of the treated group post-treatment, $\bar{y}^1_0$ is the mean value of the treated group pre-treatment, $\bar{y}^0_1$ is the mean value of the control group post-treatment and $\bar{y}^0_0$ is the mean value of the treated group pre-treatment.

By measuring the impact of the railway on the value of treated and control land or property before and after the treatment, the DID model controls for reverse causality, OVB and time-invariant unobserved heterogeneity across properties (e.g. Agostini and Palmucci, 2008; Ahlfeldt, 2013; Billings, 2011; Dubé et al., 2013; Gibbons and Machin, 2005). Nevertheless, there are some limitations to the DID. First, compared to conventional HP models, the DID requires a wider distribution (treated and control) and a larger time span (pre- and post-treatment) of consistent property data, which is not always available. Second, if land or property data is grouped to obtain the same observations before and after the treatment, this leads to a reduction in the sample size. Third, the DID assumes that the unobserved factors are time-invariant, which is a reasonable assumption if the study time span is relatively short.

In addition to the DID, a recent study by Dubé et al. (2014) estimates the effect of a commuter rail service in Canada on property values by adding a spatial dimension to the DID estimator (SDID), in an attempt to allow for spatial spillover effect. The estimated values using DID and SDID are very similar, however, although the authors suggest that results from other case studies may prove the benefits of SDID over DID.

A summary of the most widely used empirical methods is provided in Table 2-2. Comparing the three methods, we find that the HP method requires the least amount of spatially or temporally varied property data compared to the other two methods, although it requires the most comprehensive set of contextual factors to reduce potential bias. The GWR method requires a wider distribution of data spatially, while the DID requires a larger time span of data for treated and control properties. Additionally, the DID allows almost all sources of bias to be controlled for, whereas the HP and GWR methods do not control for reverse causality and require special consideration to reduce bias from other sources.

Table 2-2: A summary of the most popular empirical methods in estimating the effect of railways on land and property values (source: self-produced table)

	HP	GWR	DID
Estimate type	Global estimate	Local estimate	Global estimate
Data requirement	Large amount of contextual factors	Extensive amount of contextual factors	Small amount of contextual factors, but requires data for treated and control groups pre- and post-treatment
Suitable data structure	All types (cross-sectional, repeated cross-sectional, pseudo panel and panel)	All types (cross-sectional, repeated cross-sectional, pseudo panel and panel)	Data over a period of time (repeated cross-sectional, pseudo panel and panel)
Causal inference	Considers association but not causality	Considers association but not causality	Considers causality as well as association
Omitted Variable Bias	Can be controlled, but requires a comprehensive list of factors	Can be controlled, but requires a comprehensive list of factors	Significant reduction of OVB

Some studies have also compared the results from two or more methods. As an example, Billings (2011) who employs DID and HP methods finds that only a few estimates for the residential property data using the DID model are significant. However, all the findings using the HP model for residential properties and the DID and HP results for the commercial dataset are insignificant. Gibbons and Machin (2005) also compare results using the HP and DID models and find that the former indicate a statistically larger effect of the railway on property values. On the other hand, Agostini and Palmucci (2008) suggest a larger effect of the railway using the DID approach compared to the HP models. It is worth mentioning, however, that the data structure used in the two studies differ, which may have resulted in the difference in the magnitude of the estimates: while Agostini and Palmucci (2008) used repeated cross-sectional data, Gibbons and Machin (2005) constructed pseudo panel data. We discuss data structure in the next section.

2.2.5 Data time span and structure

The time span of property datasets ranges across studies from pre-announcement to post-operations. For example, while Agostini and Palmucci (2008) obtained property data from pre-announcement (defined as pre-treatment) to during construction, others were able to obtain data before the announcement of the transport system to a few years after operations (Billings, 2011; Concas, 2012). On the other hand, researchers like Ahlfeldt (2013) and Gibbons and Machin (2005) obtained property data post-announcement to post-operations.

In addition, the definition of the treatment time differs across case studies and mainly depends on data availability. While some researchers have assumed the year of the operations as a pre-treatment year, others have considered it as a post-treatment year. For example, although the Jubilee Line extension and the Dockland Light Rail (DLR) in London started operations in May 1999, and the full network was operational by the end of that year, Gibbons and Machin (2005) assume the year 1999 as a pre-treatment year whereas Ahlfeldt (2013) consider it as after the treatment. The data availability may have affected the choice of the treatment year in these studies; Gibbons and Machin (2005) obtain a much shorter period of repeated cross-sectional data (just five years of property data) compared to Ahlfeldt (2013), who obtains property observations from 1995 to 2008. Comparing the results from both papers, the most likely value by Gibbons and Machin (2005) suggests that a kilometre distance reduction to a rail station due to the opening of a new railway line increases the value of properties located within a 2 km radius of a station by 2.1%, whereas Ahlfeldt (2013) suggests a higher effect of 4.8%.

Another example is that of Wu (2012), who studies the effect of Beijing railway improvements on land values; the author defines the year of operations as the post-treatment year. Property observations were grouped to the years before the opening of the first two lines in 2003 (i.e. from 1999 to 2002), before the opening of four additional lines in 2008 (i.e. from 2003 to 2007) and during the construction of more lines in 2012 (i.e. in 2008 and 2009). These classifications enabled the researcher to evaluate the impact of the new railway at different stages.

On the other hand, a few researchers have relied on the start month of an event to classify property observations to pre- and post- the treatment. For example, Agostini and Palmucci (2008) examine the effect of the announcement and the start of construction of the metro on property values using transaction data for 5 and 36 months before and after the announcement, respectively, and 12 and 29 months before and after the start of construction, respectively. In summary, there seems to be a variation across studies in the cut-off point that defines the years pre- and post- a transport treatment, depending on the available data and the judgement of the researchers.

There are several types of property data that can be used to estimate the effect of a transport system on land or property values; these are cross-sectional, repeated cross-sectional, panel data and pseudo panel data. While the first three contain observed records, pseudo panel data is created from repeated cross-sectional data. Cross-sectional data is spatially distributed data at a given time, whereas repeated cross-sectional data contains multiple cross-sections over a period of time, although the same observations may not necessarily be repeated.

Cross-sectional data is problematic since it does not control for changes in the railway system over time, hence it does not allow for sources of unobserved heterogeneity. By using repeated cross-sectional observations pre- and post-treatment, and controlling for location specific effects, repeated cross-sectional data allows for the time-invariant unobserved factors across properties (e.g. Agostini and Palmucci, 2008; Billings, 2011; Dubé et al., 2013; Koster et al., 2010; Martinez and Viegas, 2009; McMillen and McDonald, 2004; Weinberger, 2001). In other words, given that at least one sale and one re-sale of the same property, or a cluster of similar properties, has occurred before and after the treatment, unobserved heterogeneity related to that property or the cluster of properties is removed.

Agostini and Palmucci (2008), for example, include a number of location control variables to account for unobserved heterogeneity across a cluster of properties. In addition, Koster et al.

(2010) use all records in repeated cross-sectional data and argue that as the sample size is large (over 55 thousand properties) and that each property is observed at least twice over a 12 year time span, using location specific effects per property observation sufficiently reduces bias. Similarly, Billings (2011) uses repeated cross-sectional data in a DID and HP context and controls for specific effects related to a cluster of similar properties.

Panel data is multi-dimensional data that contains observations for the same land or property over a period of time. In the absence of genuine panel data, some researchers have created pseudo panel data, which was first introduced by Deaton (1985). Pseudo panel data is constructed by creating cohorts that contain individual records from the repeated cross-sectional data such that a cohort repeats over time (Cameron and Trivedi, 2005; Collado, 1997; Deaton, 1985; Verbeek and Nijman, 1992). The observation in a cohort in the pseudo panel data is the mean value of the individual records in that cohort at a given time.

Compared to cross-sectional and repeated cross-sectional data, panel data is able to correct for time-invariant unobserved factors related to each property observation instead of a cluster of properties (Cameron and Trivedi, 2005; Hsiao, 2003). Although pseudo panel data also allows one to control for unobserved heterogeneity across cohorts, it is likely to contain measurement error if an insufficient number of records are grouped over time.

The previous empirical work has only estimated the effect of a transport system using one form of data structure. A few studies use cross-sectional data to estimate the effect of railways on land and property values by obtaining a month to a few months' worth of data (e.g. Al-Mosaind et al., 1993; Armstrong, 1994; Du and Mulley, 2006; Efthymiou and Antoniou, 2013; Martinez and Viegas, 2009; Pan and Zhang, 2008). The majority of studies, however, obtain repeated cross-sectional property data either for observations before and after the opening of a transport service or only after the system started operating (Agostini and Palmucci, 2008; Bowes and Ihlanfeldt, 2001; Duncan, 2008; Gibbons and Machin, 2005; Weinberger, 2001). Additionally, a very limited number of researchers obtain genuine panel data (Baum-Snow and Kahn, 2000) and, to the best of our knowledge, only two studies estimating the effect of a railway on property values have used pseudo panel data (Ahlfeldt, 2013; Gibbons and Machin, 2005).

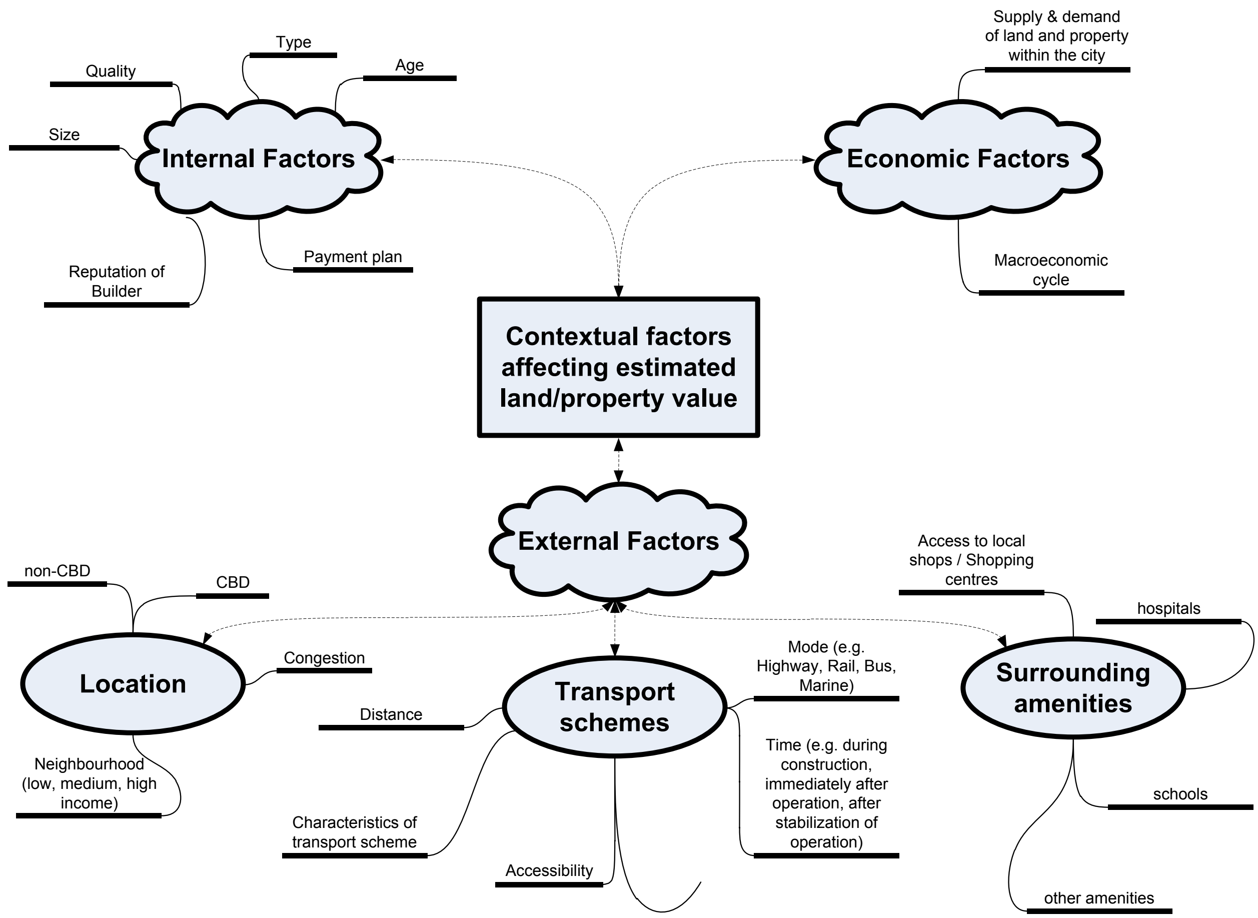
It is worth mentioning that the results can differ across data structures. For example, results using repeated cross-sectional data corrected for unobserved heterogeneity may not be similar to results using pseudo panel data, especially given that measurement error is likely in the latter dataset. Almost all case studies have considered one data structure and did not test the difference in results when using different arrangements of the data. This study tests for the difference in estimates of the effect of the metro using repeated cross-sectional data and pseudo panel data.

2.3 Factors affecting land and property values

Following a review of the literature estimating the effect of transport in general and railways in particular on land and property values (e.g. Agostini and Palmucci, 2008; Al-Mosaind et al., 1993; Armstrong, 1994; Bae et al., 2003; Clower and Weinstein, 2002; Dewees, 1976; Gibbons and Machin, 2005; Ryan, 1999; Vichiensan and Miyamoto, 2010; Weinberger, 2001), this section presents the contextual factors that affect values across case studies. This helps to identify the type of data required for such a study.

Based on the literature, we propose a classification of the factors affecting land and property values to three main categories, as shown Figure 2-2 (internal factors, external factors and economic factors). While internal factors are those related to the land use type, type of property and internal characteristics of a property that influence its value, external factors are related to the geographical location of land and property. Economic factors are city-wide factors that affect the value of land and property at large. Each is explained further below.

Figure 2-2: Factors that are expected to affect land and property values (source: self-produced figure)



2.3.1 Internal Factors

The internal characteristics of a property, such as its surface area, the number of bedrooms and property age, can have substantial effects on property values. Not all researchers, however, report the effect of these attributes when estimating the effect of railways on land or property values. For those that report coefficient values, as expected, the larger the area of a property and the number of bedrooms, the higher its value, while the results are mixed for property age (e.g. Agostini and Palmucci, 2008; Billings, 2011; Dewees, 1976; Koster et al., 2010; Martínez and Viegas, 2012; McMillen and McDonald, 2004). In addition, the value per unit area of a commercial property is generally higher than that for a residential property.

Similarly, building attributes (such as availability of parking, heating systems, a gym, and a porter service) may influence the values (e.g. Agostini and Palmucci, 2008; Dewees, 1976; Koster et al., 2010; Martínez and Viegas, 2012). Other factors that may influence the value include the reputation of the builder or the payment plan to purchase a property. In addition, there may be other specific factors to a property that are not observed but affect prices, such as the natural lighting in a property and the building quality.

2.3.2 External Factors

External factors are divided into three sub-categories: the location within the city, the surrounding amenities and the transport schemes in the vicinity. The location of land and property is one of the (if not the most) influential factor on the values. For example, accessibility to the CBD can impact the value of land and property, especially in mono-centric cities (Kobayashi and Lane, 2007). The surrounding amenities such as access to shops, schools, hospitals and other facilities that add benefits to the residents of, or commuters to, an area can also affect values (Al-Mosaind et al., 1993; Gargiulo and De Ciutis, 2010; Medda, 2008).

Accessibility to transport schemes also influences the value of land and property, but accessibility is measured differently in studies. Some define it as the time to reach a station (e.g. Pan and Zhang, 2008; Vichiensan and Miyamoto, 2010), others as the distance of a given property to a station (e.g. Agostini and Palmucci, 2008; Bae et al., 2003; Gibbons and Machin, 2005), while some use a binary value that equals one if a property is within a particular catchment area from a station and zero otherwise (e.g. Al-Mosaind et al., 1993). In addition, the characteristics of a transport system, such as the frequency of the service, the quality and the modal integration, can influence the value of land and property in the vicinity of the scheme. In some cases, the maturity of the transport system may also affect prices. Property values can be different at the time of project announcement, during construction, immediately after opening and when the system stabilizes (see Agostini and Palmucci, 2008; Concas, 2012).

The scale of the impact of a transport system on the value of land and property can differ depending on the neighbourhood type and the value of the transport scheme to the residents or employers of an area. For example, some studies have suggested a larger positive impact on the value of land and property located in dense, congested and CBD areas (e.g. Cervero, 2003; Duncan, 2008). Others indicate a larger uplift in the value of properties sited in low income neighbourhoods, since accessibility to a public transport service tends to be valued more in these neighbourhoods (Diaz, 1999; Medda, 2008). In general, the impact of a new or an improved transport system on land and property values is greater at locations where the new system offers an enhanced accessibility benefit (Edwards and Cooper, 2009; Mikelbank, 2004).

2.3.3 Economic Factors

The value of land and property may vary with the macroeconomic or the business cycle in the city or due to changes in the supply and demand of land and properties (Ryan, 1999). The macroeconomic cycle is a full fiscal cycle that reaches a high and a low peak over a period of time, and is affected by various economic factors such as the gross domestic product (GDP) and employment rates. To control for the effect of economic factors on land or property values, studies aim to obtain land or property data over a period of time and to add time variables into the estimation models as well as relevant contextual factors like employment density.

The supply and demand of land and property may not only vary across cities, but also within different areas in a city. This in turn can impact values overall or at certain neighbourhoods. The supply and demand is closely related to the regulations and the planning parameters imposed in the city. For example, government decisions to build more properties, allow for the growth of developments or sell more land affect the volume of the available land or property in the market. In addition, legislation on land or property ownership, property tax and mortgage schemes are examples of the regulations that affect consumers' choices on purchasing, leasing or renting land or properties. Nonetheless, and mainly due to insufficient or lack of available data, the effect of supply and demand is rarely captured in studies estimating the effect of a transport system on property values.

2.4 Fiscal mechanisms to capture the increase in land and property values due to investments in railways

Since many studies have suggested a positive effect on the value of land and property due to their accessibility to railway stations, some researchers have argued that partially capturing the monetary value of such increases can be justified as a means of funding at least a part of the transport system (Dickens, 1992; Smith and Gihring, 2006). In fact, a number of studies estimating these impacts have this objective in mind (Cervero and Susantono, 1999; Du and Mulley, 2006). This study has, in part, the aim of exploring the policy implications of the estimated effect of the Dubai Metro on property values in the context of value capture, and this section summarises some fiscal mechanisms that might achieve this. These are: taxes or charges, partnership deals and endowment schemes (see Enoch et al., 2005; Lari et al., 2009; Martínez and Viegas, 2012; Medda, 2012; Salon and Shewmake, 2012).

The first method (taxes or charges) relies on capturing continuous payments though taxes or levies on land and property that benefit from accessibility to transport services. The advantages and disadvantages of this fiscal mechanism are as follows. First, while in the majority of cases taxes and charges allow for a continuous payment towards the transport service, the collected cash is generally limited. Second, in cities that impose taxes, charges on land value can be justified and an accepted value capture mechanism. Finally, this funding scheme is flexible in terms of value capture and can be applied across cities at different rates if needed.

Partnership deals rely on partnering with developers or landowners to develop the area around a transport scheme and use part of the profits to fund the service. The main advantages of partnership deals are the following. First, the capture value is generally higher than that from taxes. Second, the risks of the development around the transport scheme are transferred from the government to the developer or landowner. Third, there is flexibility regarding the payment plan since the agreement can stipulate whether it is a one-off payment at the time of the agreement or a continuous payment over a number of years. There are two main limitations of this mechanism, however. First, such a scheme assumes that land is available around the transport scheme for developers. Second, it is more likely to succeed in areas where the benefits of the transport scheme are clearly evident.

There are various forms of endowment schemes. In some, governments or transport providers develop the land in close proximity to railway stations and use the money to fund the transport scheme. In other cases, the transport provider is granted land by the government near railway stations to resell at a later date after its value increases due to the planned or constructed transport system. The cash is then used to support the costs. The advantages and disadvantages of endowment schemes are very similar to partnership deals. First, the capture value is generally higher than that with a fiscal mechanism. Second, there is also flexibility on the payment plan, however it depends on the type of the development; the scheme can be a one-off payment if the land or property is resold after it gains value, and it can be continuous if land is developed to generate revenue over a number of years. In contrast to the partnership deals, the risks remain with the transport provider. In addition, the two limitations are similar to the partnership deals. First, this scheme assumes that empty land is available around the transport scheme and, second, it is more likely to succeed in areas where the benefits of the transport scheme are clearly evident.

2.5 Conclusions

This chapter has provided a review of the recent literature related to the impact of railways on land and property values. Starting with the theory that explains the interaction between access to transport systems and land and property values, it has examined the various types of relationship between the values of land and property and their distance from a transport system.

The chapter then reviewed estimates of the effect of railways on land and property values from around the world. Although there is no objective research estimating the effect of the Dubai

Metro on land or property values, a wide variation of empirical evidence was observed across other case studies. For example, the magnitude and the radius of impact varies with land or property type (e.g. land versus property, residential versus commercial) as well as the geographical location, both globally and within the study area (e.g. Europe versus North America, CBD versus non-CBD). In addition, potential sources of bias were considered and the empirical methods and data structures used in the literature were discussed.

Next, the review of the literature identified the contextual factors that affect land and property values. These could be classified into three main categories (internal, external and economic factors), with a brief explanation offered for each. Finally, since one of the motivations for studying the impact of railways on land or property values is to capture part of the value increase due to the railway, this chapter reviews three main fiscal mechanisms applied globally. This discussion aims to set out the background for the discussion of the policy implications of the study results in the context of value capture (in chapter 10).

The above review and discussion of the differences across case studies in estimating the effect of railways on land and property values is based on a subjective summary and interpretation of the existing literature. There is a need also to analyse objectively the sources of variation across studies to identify the most influential factors in determining the effect of railways on land or property values. A meta-analysis in this regard is conducted in the next chapter.

Chapter 3. A META-ANALYSIS OF THE IMPACT OF RAIL ON LAND AND PROPERTY VALUES

3.1 Introduction

As explained in the literature review chapter, studies indicate a wide range of empirical evidence for the impact of railways on land and property values. The majority of studies suggest that accessibility to railway stations enhances property values (e.g. Agostini and Palmucci, 2008; Laakso, 1992; Pan and Zhang, 2008; Voith, 1991), however the range of the impact across case studies is large (Mohammad et al., 2013). Some, however, indicate a negative (e.g. Du and Mulley, 2007) or no significant effect of proximity to rail stations (e.g. Clower and Weinstein, 2002).

This chapter studies the large variation in these assessments of the impact of railways on land and property values in an attempt to understand the reasons for these differences.[5] Although a few studies have attempted to explain the large range of estimates, the majority have not applied an objective approach to their analyses (e.g. RICS Policy Unit, 2002; Zhang, 2009) or have analysed only a limited number of factors. Ryan (1999), for example, considered only the accessibility measure (i.e. time-based and distance-based accessibility) to examine the difference in estimates for the effect of railways on land and property values. The only objective study of the variation in impact is the meta-analysis by Debrezion et al. (2007) which includes a sub-set of potential study-design factors; these are property type, type of public transport system, empirical method, data collection time period, accessibility to roads and demographic variables.

This study contributes to the existing literature in the following ways. First, while Debrezion et al. (2007) concentrated on railways in the United States of America (USA) and gathered 57 estimates, the meta-analysis presented here analysed the variation across three continents using

[5] The work of this chapter has been published in Transportation Research Part A – Policy and Practice (Mohammed et al., 2013)

102 observations from 23 studies. The estimates explore the effect of railways on land and property values in Europe, Asia and North America between 1980 and 2007 and suggest a significant geographical variation in results.

Second, this meta-analysis considers a more comprehensive set of contextual and methodological factors that could affect land and property values than the factors used by Debrezion et al. (2007). In more detail, fifteen contextual factors are either added or modified and five new methodological factors are added, with some of these being found to be significant. Third, researchers may refer to this study for an indication of the effect of railways on land and property values, especially in the absence of data or estimates for a given railway. Fourth, for the first time in the literature in this area, a check was performed for publication bias in order to test whether only positive and significant results are reported.

This chapter is structured as follows. The next section introduces and defines the scope of the meta-analysis. The design of the empirical model and the attributes that are expected to affect the variation in results are discussed in section 3.3. Section 3.4 presents the results of the meta-analysis and the conclusions are in section 3.5.

3.2 Scope of the meta-analysis

Meta-analysis allows the objective and systematic analysis of variation across findings using regression methods (e.g. Melo et al., 2009; Stanley and Jarrell, 2006). This approach provides a fair basis for comparison provided that one understands and uses the rules that govern the analysis. First, in order to compare similar empirical results across studies, the dependent variable needs to be unified (DeCoster, 2004). Studies report the estimates of the effect of a railway on land and property values differently; while the majority of researchers report a percentage change, others present the change as a monetary value. Standardization of the dependent variable is not always feasible, however, since the amount of information in some case studies may not be appropriate for the conversion. Here, the percentage change in land and property values was chosen as the unit for the dependent variable, with the monetary values being converted by comparing the change in land or property value within the vicinity of the railway to the change further away.

Second, the number of estimates per study used in the meta-analysis can affect the results. On the one hand, choosing one empirical result per study can bias the meta-analysis to the choice of that estimate. On the other hand, using more than one finding can lead to a correlation between estimates. In this study, we choose more than one estimate per case study in order to increase efficiency due to the increase in the number of observations and we also use a random-effects estimator to account for within- and between-variations of estimates.

Third, using results from published articles only may bias the estimates to positive and statistically significant values. In this study, therefore, both published and unpublished studies are used. Fourth, similar to any regression model, the contextual and methodological factors used affect the results. Since there is no particular direction in the literature to the choice of factors for the meta-analysis, the most appropriate factors chosen for this study are based on the review of previous empirical work provided in the previous chapter.

A summary of the meta-sample is shown in

Table 2-1 and histograms of the variation in property values across studies in appendix A. The histograms show the significant estimates in the considered meta-sample and present the percentage change in the value of land or property due to a railway. It is found that the mean effect of railways is 0.08 with a standard deviation of 0.172. The range of variation in land and property values is quite large (from -45% to above 100%).

3.3 Design of the meta-analysis

The sources of variation in the estimates for the effect of railways on land and property values can be related to the context of the study (contextual factors) and the employed estimation method (methodological factors). Contextual factors are the parameters that affect land and property values in reality, whereas methodological factors include the analysis type and the data structure (Table 3-1). The table also specifies which factors are new and which were considered in the previous meta-analysis by Debrezion et al. (2007).

The variables considered in this study are as follows. We distinguish the effect of railways on land versus property values by specifying a value of 1 for results on property values and zero otherwise. Since the impact of railways on the value of a property listed for sale compared to that for rent may also differ, a dummy variable 1 is specified for the estimates on the purchase price. Similarly, to test for the variation in results due to the type of property (residential, retail, office) dummy variables are used for each estimate type, keeping residential land and property as the reference case. In addition, studies that have controlled for property characteristics in their models are distinguished from those that have not. Another empirical dimension considered in the meta-analysis is time. The reported effects on land and property values in both the 1990s and 2000s are compared to the estimated impact in the 1980s.

A set of control variables is also included to test for the variation in estimates due to the railway type (light rail, heavy rail or metro and commuter rail).[6] Each case study specifies the type of the considered railway and the same is used to classify the railways in this meta-analysis. In addition, we expect that the maturity of the railway may influence the magnitude of the impact on land and property values, therefore we specify a set of control variables for the results reported at four stages (at the announcement of the rail project, during construction, within a year of operations beginning (i.e. immediately after operations begin) and at more than a year after operations begin (i.e. at stabilization of the rail service). Since the literature suggests that the effect of railways on the value of land and property can differ according to the distance to stations (e.g. Al-Mosaind et al., 1993; Billings, 2011; Bollinger et al., 1998; Bowes and Ihlanfeldt, 2001; Du and Mulley, 2006), dummy variables are included for three distance bands (zero to 200m, 201m to 500m and 501m to 806m) compared to a threshold of more than half a mile (above 806m) from a station.

One of the main aims of this meta-analysis is to examine the variation in estimates across continents. The influence of railways on land and property values in Europe, East Asia, other parts of Asia and North America are controlled for. This particular geographical classification is based on the transport modal share (private car versus public transport) in the cities located in each geographical zone. In addition, since the competition between railways and private transport services may affect the impact of railways on land and property values, the accessibility of land and property to roads is controlled for as well.

Another dimension in the meta-model is the location of land or property within a city (i.e. in a CBD versus a non-CBD area). Most studies report an average effect of railways on the value of land or property in all parts of the city, here, however, a dummy variable of 1 is used if the study reports an estimate for land or property in either all areas or just a non-CBD area and zero if the

[6] Although we use the same definition of the railway type as indicated in each case study, here is a brief on each type. Light rail, heavy rail and the metro are railway systems that run within a city, however in general, light rail carries fewer passengers and runs at a lower speed. Commuter rail, on the other hand, carries the largest number of passengers and connects different cities or towns.

estimate is for a CBD area only. The final contextual factor that is controlled for is whether the study considers the effect of neighbourhood characteristics on land and property values in the empirical models.

Moving to the methodological factors, a set of control variables are specified to examine the effect of data type (cross-sectional versus panel or time-series data) on the empirical results. In addition, the method used in a study may affect the reported estimates and is therefore controlled for in the meta-analysis. The difference in the estimates due to three approaches (geographically weighted regression (GWR), difference-in-differences (DID) and simple comparison of average value changes) is tested in comparison to the results obtained using hedonic pricing (HP) models. A set of dummy variables are also included to compare the effect of railways on land and property values due to the model type (a semi-log and a double-log model compared to a linear model). Finally, the analysis distinguishes the rate of the predicted impact of railways based on the statistical significance of the estimate.

A common regression model for meta-analysis is shown below (see also Melo et al., 2009; Stanley and Jarrell, 2006; Weichselbaumer and Winter-Ebmer, 2005); a semi-log model is chosen for the meta-analysis in tandem with the majority of models estimating the effect of railways on land and property values.

$$\ln P_{ij} = \alpha_0 + \sum_k \alpha_k D_{ij,k} + \mu_j + \varepsilon_{ij} \quad (4)$$

where,

lnP_{ij} is the logarithm estimate of land or property value change for the individual estimate(s) i obtained from a given study j

$D_{ij,k}$ is the meta-regressor k

μ_j is a measure of study-specific effect

ε_{ij} is the model disturbance term.

Similar to other regression models, one may use a fixed-effects (FE) or a random-effects (RE) estimator in the meta-model. We find that a RE estimator is preferable for the following reasons. First, a heterogeneity test[7] (Q) is used to identify if the actual change in land or property values due to railways varies significantly across studies: a substantial value of Q implies that each predicted impact of railways on the value of land or property is an estimate of a true impact that is statistically different for each railway. Since we obtain a statistically significant value of Q, the RE estimator is suitable for this meta-analysis. Second, since more than one estimate per study is included in the meta-sample, there is probably a correlation between the empirical results from the same study and the RE estimator accounts for variations within a study and between studies.

Third, it can be contended that the meta-sample is drawn from a random population of available reports, and this is explained as follows. As many studies as possible were collected from various sources (journal articles, conference proceedings, research working papers and online reports), either by searching online in journal websites for reports on the effect of railways on land or property values or by searching for particular papers referenced in the previous empirical work. Some material was not available online and was obtained from the hard copy versions of journals from the college's library. In addition, Google was used to search for unpublished work and online research papers.

Nevertheless, the meta-sample does not cover all available studies: some reports either lacked some of the factors considered in this study or reported an estimate for the effect of railways on land or property values that was not compatible with the dependent variable used in this study (a percentage change in value) without sufficient information available to convert it. The obtained reports, however, cover case studies from over 80 cities located in three continents over a period of just under 30 years.

[7] The result for the value of Q equals 4.2×10^9 and it is significant at 1%.

A number of contextual and methodological factors that affect land and property values are controlled for, and examine the differences in results due to these factors are examined. Since the meta-analysis aims objectively to analyse the variation across estimates for the effect of railways on land and property values, estimates from different case studies are considered. Additionally, when at least one of the contextual or methodological factors considered in a given study varies, it was decided to add more than one estimate per study. The meta-analysis, therefore, allows us to identify the average differences between estimates. In other words, it does not necessarily identify the average differences in results across case studies or within a study, but rather the average differences across estimates.

Table 3-1: Regressors used in the meta-analysis

Dimension	Variable	Definition	Reference Case
Contextual factors			
Property or land	D_{PP}*	1 if the outcome is a measure of property value, 0 otherwise	Study reports on land values
Reported value	D_{P}*	1 if purchase or sales price, 0 otherwise	Study reports rent values
Type of land or property	D_{comm}	1 if land or property type is retail, 0 otherwise	Land or property type is residential
	D_{off}*	1 if land or property type is office, 0 otherwise	
Property characteristics	D_{pc}	1 if property characteristics are used in study analysis, 0 otherwise	Study does not report on property characteristics
Time of case study data collection	$D_{90\text{-}99}$**	1 if the data time is between 1990 and 1999, 0 otherwise	Data is for the time period from 1980 to 1989
	$D_{00\text{-}10}$**	1 if the data time is between 2000 and 2010, 0 otherwise	
Type of rail service	D_{HV}	1 if heavy rail or metro, 0 otherwise	Rail service is light rail (LRT)
	D_{commu}	1 if commuter rail, 0 otherwise	
Rail system maturity	D_{cons}*	1 if the data are obtained during construction of rail service, 0 otherwise	Data obtained within a few months after project announcement
	D_{imm}*	1 if the data are obtained immediately after operation of service, 0 otherwise	
	D_{stab}*	1 if the data are obtained after system stabilization, 0 otherwise	
Distance to rail station	$D_{0\text{-}200}$**	1 if distance to rail station is between 0-200m, 0 otherwise	Distance to rail station is more than 806m (i.e. half a mile)
	$D_{201\text{-}500}$**	1 if distance to rail station is between 201-500m, 0 otherwise	
	$D_{501\text{-}805}$**	1 if distance to rail station is between 501-805m, 0 otherwise	
Geographical location	D_{EU}*	1 if the study is in Europe, 0 otherwise	Study is in the North American cities
	D_{EA}*	1 if the study is in East Asia, 0 otherwise	
	D_{A}*	1 if the study is in West Asian cities, 0 otherwise	
Accessibility	D_{acc}	1 if the study uses accessibility to roads in the analysis, 0 otherwise	Study uses accessibility to rail stations only
Land or property location	$D_{non\text{-}CBD}$*	1 if land or property is either not in CBD or in both CBD and non-CBD, 0 otherwise	Land or property is in CBD area
Neighbourhood type	D_{nc}	1 if neighbourhood characteristics are used in study analysis, 0 otherwise	Study does not report on neighbourhood characteristics
Methodological factors			
Data type	D_{cs}*	1 if the study uses cross sectional data, 0 otherwise	Study uses panel or time-series data
Analysis method	D_{GWR} *	1 if the study uses geographically weighted regression, 0 otherwise	Study uses Hedonic Price model
	D_{DID} *	1 if the study uses difference-in-difference model, 0 otherwise	
	D_{comp}*	1 if the study compares average value changes over time, 0 otherwise	
Model type	D_{SL}	1 if the model type is semi-log, 0 otherwise	Model type is linear regression
	D_{DL}	1 if the model type is double-log, 0 otherwise	
Results type	$D_{non\text{-}sig}$*	1 if result type is not-significant, 0 otherwise	Result type is significant

Legend: * These variables are not used in the study by Debrezion et al. (2007).
** These variables were used in the study by Debrezion et al. (2007), however not to the level of detail as those here.

3.4 Results

Two sets of meta-analysis models are considered with the results being reported in Table 3-2. The first model tests the percentage change in land and property values following the development or improvement of a railway in relation to all of the factors listed in Table 3-1. The second model, on the other hand, tests for the effect of the most influential variables (internal factors to land and property, and transport related factors) on the reported land or property values.

Models 1 and 2 returned comparable results, which implies a large influence of internal and transport related factors on the value of land or property. In addition, there was a larger overall goodness of fit value for model 1 (60%) than for model 2 (46%). This is to be expected, since there is a more comprehensive set of variables in the former model and the additional factors substantially explain some of the differences in the effect of railways on land and property values across case studies. The findings of models 1 and 2 are discussed in more detail in this section.

Starting with the contextual factors, the results from models 1 and 2 indicate that the effect of rail on the value of properties is substantially lower than the effect on land values by 22.3 and 16.3 percentage points, respectively.[8] Previous studies have suggested that almost all land experiences an increase in value due to accessibility to railway stations, whereas properties experience both a rise and a reduction in values depending on the area. A possible interpretation of this finding is that while empty land can potentially allow for a more varied development scheme taking advantage of improved accessibility to the rail network, property that has already been built is fixed and the impact of increased accessibility to that property is therefore less marked. In addition, the findings from both models imply no noticeable difference between the effect of railways on the sale value versus the rental value.

[8] Since the dependent variable in our model is the percentage change in land or property value, the unit of the coefficients is the percentage point.

Table 3-2: Meta-analysis results

Factor type	Dimension	Variable	Model 1		Model 2	
	Constant	α_0	0.5759 (0.2652)	**	0.3652 (0.2246)	*
Contextual – internal factors	Property or land	D_{PP}	-0.2230 (0.0862)	***	-0.1626 (0.0813)	**
	Reported value	D_P	0.1167 (0.1045)		-0.0329 (0.0920)	
	Type of land or property	D_{comm}	0.3156 (0.0581)	***	0.2431 (0.0530)	***
		D_{off}	0.1342 (0.0909)		0.0211 (0.0886)	
	Property characteristics	D_{pc}	-0.0918 (0.1823)		-0.0172 (0.1727)	
Contextual -time	Time of case study data collection	$D_{90\text{-}99}$	0.0311 (0.0731)			
		$D_{00\text{-}10}$	0.0855 (0.0625)			
Contextual – External - Transport scheme	Type of rail service	D_{HV}	-0.0488 (0.0693)		-0.1165 (0.0513)	**
		D_{commu}	0.2531 (0.0600)	***	0.2431 (0.0586)	***
	Rail system maturity	D_{cons}	-0.0175 (0.0827)		0.0309 (0.0796)	
		D_{imm}	-0.0369 (0.0859)		-0.0949 (0.0874)	
		D_{stab}	-0.1452 (0.0879)	*	-0.1843 (0.0741)	**
	Distance to rail station	$D_{0\text{-}200}$	0.0749 (0.0613)		0.0797 (0.0654)	
		$D_{201\text{-}500}$	0.0706 (0.0441)		0.0627 (0.0458)	
		$D_{501\text{-}805}$	0.0872 (0.0525)	***	0.0946 (0.0515)	*
	Geographical Location	D_{EU}	0.1486 (0.0591)	**	0.0935 (0.0487)	*
		D_{EA}	0.1575 (0.0754)	**	0.0230 (0.0637)	
		D_A	0.1231 (0.1210)		0.0606 (0.1198)	
	Accessibility	D_{acc}	-0.1491 (0.0666)	**	-0.0279 (0.0487)	
Contextual – external – Location	Land or property location	$D_{non\text{-}CBD}$	-0.0141 (0.0885)			
	Neighbourhood type	D_{nc}	-0.0441 (0.1042)			
Methodological – Data type, analysis method and model type	Data type	D_{cs}	-0.1598 (0.0551)	***		
	Analysis method	D_{GWR}	-0.0828 (0.0818)			
		D_{DID}	-0.0156 (0.0964)			
		D_{comp}	-0.3202 (0.1410)	**		
	Model type	D_{SL}	-0.1498 (0.0661)	**		
		D_{DL}	-0.1830 (0.1103)	*		
	Results type	$D_{non\text{-}sig}$	-0.0102 (0.0503)			
	Observations		102		102	
	R^2 (total)		0.6028		0.4647	
	R^2 (within)		0.4322		0.2752	
	R^2 (between)		0.8024		0.7426	

Legend: *, **, *** indicate significance at 10%, 5% and 1% respectively. Values in parentheses are standard errors.

The meta-analysis reveals that while proximity to a rail station enhances the value of retail properties much more than it does for residential properties (by about 31 and 24 percentage points for models 1 and 2, respectively), no significant difference is observed for the effect on the value of offices versus residential properties. It is no surprise to observe a higher impact on the value of retail properties, since the literature also indicates that the potential benefit to retail property of proximity to a railway station is generally higher than for other land uses (Debrezion et al., 2007). As for the result on office values, giving that commuting trips to work are often regular, the value of accessibility from a railway to dwellings may be considered similar to the value of accessibility to offices.

In addition, the findings indicate that controlling for property characteristics in the estimation model does not impact the variation in estimates for the effect of railways on land and property values. This is expected, since the physical attributes of a property are not related to the location of railway stations. In addition, the coefficients on the time of the reported estimate (model 1) indicate that the range of variation in the estimated effect of railways on land and property values across decades is not statistically different.

This study has also examined the difference in the accessibility benefit in respect to different railway types. The results from model 1 indicate no noticeable difference between the effect of metro or heavy rail compared to the effect of light rail on land and property values, whereas model 2 suggests that the value of land and property near a metro or heavy rail station is reduced by 12 percentage points. This finding is not surprising since the negative externalities (e.g. an increase in noise and pollution levels) due to proximity to a metro or heavy rail station is larger than that for light rail. Nevertheless, models 1 and 2 indicate that proximity to a commuter railway increases the value of land and property by 25 and 24 percentage points, respectively, compared to the value of land and property in the vicinity of a light rail station. The studies done by Cervero and Duncan (2002) and Weinstein and Clower (1999) also confirm this result. A possible explanation is the greater benefit of commuter rail compared to a light railway in terms of providing access to the wider rail network for long-distance trips.

The controls on the maturity of the rail service suggest no significant differences across studies at different stages of the rail system except after the system stabilizes; the findings from models 1 and 2 indicate that the effect of railways is on average 15 and 18 percentage points lower, respectively, compared to the “at announcement” stage. One reason for this is that the perceived benefit of accessibility to a railway station at announcement can be higher than the capitalized benefit after the system stabilizes (e.g. Bae et al., 2003).

Observing the results across the catchment zones, no significant difference is found in the impact of railways on the value of land and property located within 500m of a station compared to those located beyond half a mile. One explanation for this is the possible impact of negative externalities in close proximity to a railway station (e.g. increase in noise and crime levels) in some cities, which in turn results in a similar effect of the railway on the value of land and property within 500m compared to those beyond 806m. Moving to the impact on the value of land and property located above 500m and up to 806m of a station, the results suggest a significantly higher effect (by 9 percentage points) compared to the effect on properties located beyond half a mile.

The coefficients for the geographical location dummy variables suggest that the change in the value of land and property located in Europe (models 1 and 2) and East Asia (model 1) is significantly higher compared to the change in the value of land and property in North America, by at least 9 percentage points. This result is in line with the modal share characteristics in that the trip share using public transport services is higher in Europe and East Asia and therefore the value of accessibility to rail services is also higher. Nevertheless, no significant difference was found between the changes in land and property values in West Asian cities compared to the North American cities.

While the findings from model 1 indicate that accessibility to roads reduces the effect on the value of land and property of accessibility to a railway station by 15 percentage points, no significant difference is found in the parallel analysis using model 2. Since road transport competes with rail services in the majority of cases, it is expected that accessibility to roads would reduce the perceived benefit associated with accessibility to a railway station.

Other controls added in the meta-analysis are the location of land and property within the city, as well as whether the study included the neighbourhood type in the empirical model. No noticeable difference was found for the estimated change in the value of land and property located only in the CBD, and controlling for neighbourhood characteristics did not influence the estimated impact of railways.

The results from the meta-analysis also indicate that the choice of the methodological factors affects the estimated change in land and property values. It can be seen that the findings from studies using cross-sectional data are lower by 16 percentage points compared to the results of studies using panel or time-series data. The findings also suggest that the estimates from different types of model are similar in size, except when comparing average values. The results indicate that a comparison of the average change in the value of land and property produces lower estimates of the change, by 32 percentage points, compared to the results using HP models. This result is not expected since the HP models control for other factors and hence results would be expected to be lower. A possible reason for this finding, however, could be the limited number of studies in the meta-sample that used average values and the possibility that this small sample sized has skewed the results.

In addition, the controls for the model type imply that the estimates using semi-log and double-log models are lower by 15 and 18 percentage points, respectively, compared to the estimates from linear models. A possible explanation is the exponential form of the log models which results in a greater reduction in values compared to the linear models. Comparing this with the results of previous empirical work, Weinberger (2001) estimated the effect of proximity to a light railway station on the value of offices in Santa Clara (USA) and found that the results using a semi-log model were lower than the results using a linear model. It is worth mentioning, though, that the majority of studies have showed that semi-log and double-log models are more suitable for studies on the effect of railways on land and property values (e.g. Bowes and Ihlanfeldt, 2001; Weinberger, 2001).

Finally, the findings from the meta-analysis indicate no noticeable difference in the significant and the insignificant estimates, which in turn implies that the range of magnitudes of the impact of railways on land and property values is similar for significant and insignificant results. Mohammad et al. (2013) have also tested for publication bias and found that researchers report both positive and negative results but they tend to be biased towards statistically significant estimates.

3.5 Conclusions

This chapter has presented the results of a meta-analysis allowing an objective assessment of the variation in existing evidence on the effect of railways on land or property values. Two models were tested: model 1 included a full set of contextual and methodological factors that were expected to explain a part of the variation in results, whereas model 2 considered a subset of the most influential variables. The findings from the two models are comparable and stress the substantial influence of internal factors to land and property and the transport accessibility factors in explaining the largest part of the variation across estimates.

The results confirm that the change in land values due to proximity to railways is higher compared to the change in property values and that no significant difference is observed for the results on sale values versus rent values. While the findings imply that the effect of railways is higher on the value of retail properties compared to residential properties, no noticeable difference is found regarding the impact on offices compared to dwellings. In addition, the results suggest that controlling for property characteristics in the estimation model has no substantial influence on the variation across the estimates and that the range of variation is not significantly different over three decades.

The results also indicate that land or property in close proximity to commuter rail services experiences an increase in value, whereas land or property close to a metro or heavy rail station reduces in value compared to land and property near a light rail station. The findings from the meta-analysis imply that the change in land and property values were similar over different stages of maturity of the rail system, except for after system stabilization, where the results indicate a reduction in value compared to the values at announcement. The findings also suggest that land or property located within 500 and 806 metres of a station exhibits a greater increase in values compared to the value of land or property located beyond 806 metres.

Comparing the change in land and property values due to investments in railways across continents, it was observed that the change was more pronounced in public transport oriented cities (i.e. in Europe and East Asia compared to North American cities). Not surprisingly, the results also reveal that accessibility to roads lowers the impact of accessibility to a railway station. In addition, no noticeable difference was found in the value of land and property with regards to the location within the city (i.e. in a CBD or a non-CBD area) nor when the neighbourhood type was included in the estimation model.

Moving to the impact of the methodological factors on the variation in empirical results, the findings suggest that the estimates obtained using panel or time-series data are lower than the estimates using cross-sectional data. In addition, it was found that the results across different empirical models were similar, except when comparing average values which, surprisingly, gave lower predictions than the results from HP models. Finally, the results revealed that the reported estimates using semi-log and double-log models were statistically lower than the results obtained using linear models.

The findings from this meta-analysis allow one to consider objectively the factors that are more influential on estimating the effect of railways on the value of land or properties. As a consequence, in turning now to estimate the effect of the Dubai Metro on property values, the most significant factors, as identified in this meta-analysis, are considered (wherever data is available). In addition, the findings from the meta-analysis are considered when discussing the results from this study, as can be seen especially in chapter 10. The most relevant factors for this study are explained further below.

First, as it is found that the effect on retail properties is significantly larger than the effect on residential properties, the impact of the Dubai Metro on retail versus residential properties is estimated in separate models for each type, and the estimated effect of the metro between the two property types is also compared. Second, since the meta-analysis finds that accessibility to roads statistically affects the estimated impact of railways on land and property values, the distance to highways is included in the subsequent models.

Third, the meta-analysis reveals that the impact of railways on the value of land or property in public transport oriented regions is higher than that in car-oriented regions. Since Dubai is a car-oriented city, the effect of the Dubai Metro on the value of properties is likely to be lower than the effect of accessibility to railways in cities with relatively high public transport trip rates (e.g. European cities). Overall, therefore, the meta-analysis performed here provides a consistent basis through which to compare the estimates from this study with the findings from the previous empirical work.

Chapter 4. DUBAI IN CONTEXT

This chapter starts with a brief overview of the developments in Dubai until the year when the property data used in this study ends (2011) (section 4.1). This is followed, section 4.2, by a description of the socio-demographic characteristics of the communities in Dubai and, in section 4.3, a description of the emirate's transport system and the travel patterns within Dubai and to neighbouring emirates. The Dubai Metro (the transport innovation) is introduced in section 4.4. Final remarks from this chapter and the link to the results of this study are provided in section 4.5.

4.1 The development of Dubai

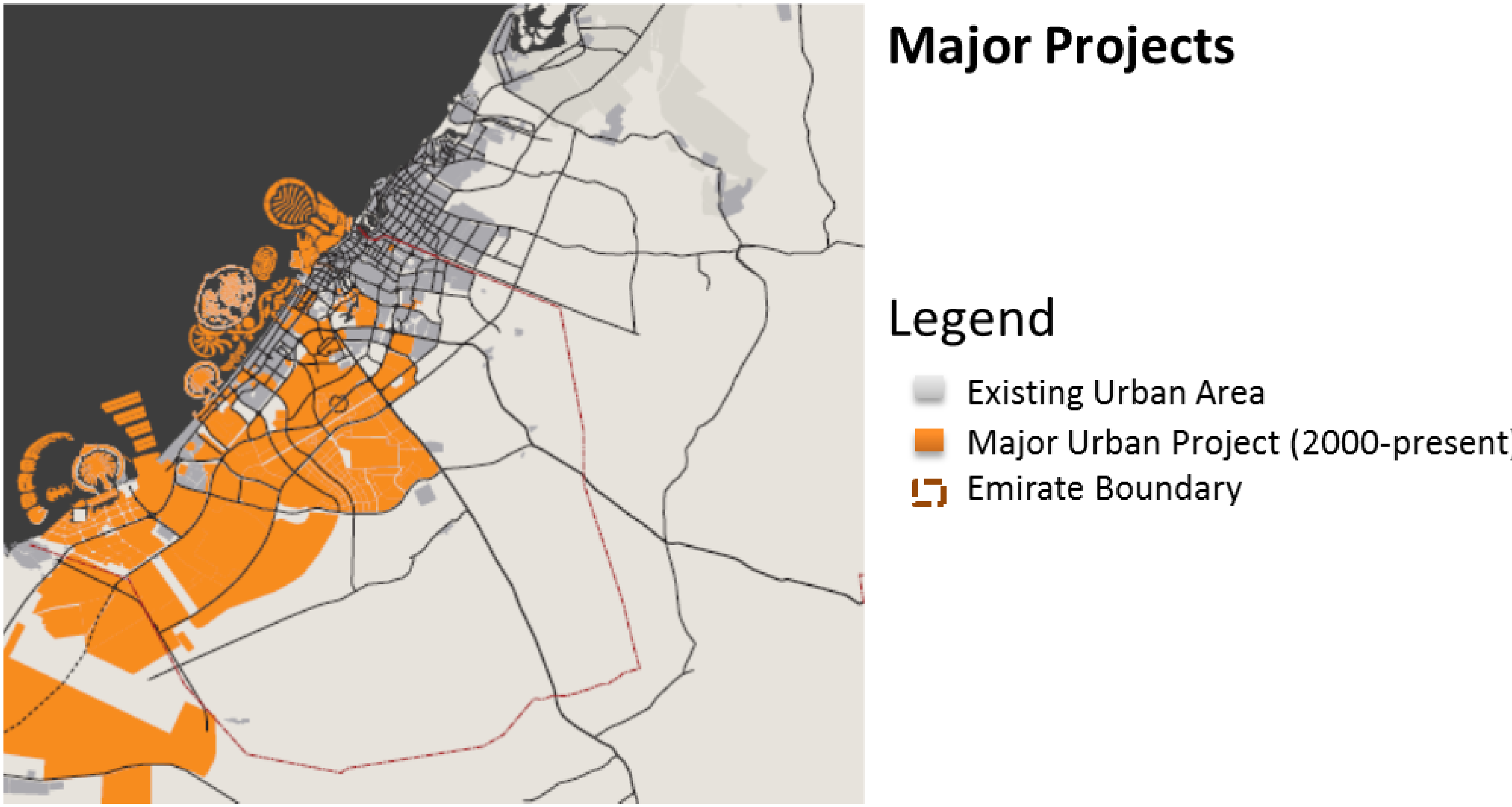
Dubai is one of the emirates in the United Arab Emirates (established in December 1971). Since the year 2000, the government has been granting large areas of empty land to local developers to plan and develop with the aim of transforming Dubai into a global city. Since then, it has been growing at a substantial rate.

Figure 4-1 presents the urban area in Dubai in the year 2000 and the allocated land for mega-projects as of the year 2010. In 2010, only about 8% of the allocated land had been developed (Roads and Transport Authority, 2010a). Examples of mega projects are the Palm Jumeirah, Downtown Dubai, Dubai Marina, Dubai Land and Dubai International City.[10] Some of these projects are on the scale of a city, such as Dubai Land, which has a population capacity of over 1.7 million.

While Dubai Municipality (DM) is responsible for the urban planning of the city, local developers are allowed by law to obtain approvals directly from the government for planning their projects (Dubai Municipality, 2010). This allows for a much quicker process to develop the land in line with Dubai's plans for rapid growth. Nevertheless, other stakeholders also play an important role in providing the supporting infrastructure to these developments, such as the Dubai Municipality, Roads and Transport Authority (RTA) and Dubai Water and Electricity Authority (DEWA).

[10] Nakheel, Emaar and Dubai Land are some of the main large-scale developers in Dubai. More information can be obtained from the developers' websites; examples are www.nakheel.ae, www.emaar.ae and www.dubailand.ae

Figure 4-1: The old urban area in the year 2000 and the location of the mega-projects that were announced until the year 2010 (source: Dubai Municipality, 2010)



The majority of developers have converted the land to mixed uses. In fact, these mega projects have together served to transform Dubai into a polycentric city, with dramatic increases in the number of commercial and residential properties. As a consequence, the population has grown from around 827,000 in the year 2000 (Cooper, 2001) to over 1.9 million in the year 2011 (Dubai Statistics Center, 2012). Employment has also grown from over 570,000 employees in the year 2000 (Dubai Statistics Center, 2002) to over 1.3 million in the year 2011 (Dubai Statistics Center, 2012). Dubai has a large mix of population groups and around 90% of residents are non-locals (Dubai Statistics Center, 2011b).

In addition, the growth in Dubai has led to growth in the neighbouring emirates of Abu Dhabi, Sharjah and Ajman, especially in terms of housing supply (Dubai Municipality, 2010). In fact, the majority of the 800,000 commuters who travel to Dubai daily are employees who live outside of the emirate, although a proportion visit Dubai for leisure or other purposes. As a result, the daytime population in Dubai reached 2.7 million in the year 2011. With two main access points (east and west), the commuters from the east side of Dubai account for around 22% of the total trips in the emirate, whereas commuters from the west side are much more limited, counting for only 2.8% (Roads and Transport Authority, 2012a).

Although the global economic recession from the end of the year 2008 affected Dubai, the majority of projects remained in place, albeit with some restructuring and rescheduling. There has, however, been less demand for properties during that time and there is evidence of reduced property prices (Dubai Municipality, 2010).

4.2 Communities in Dubai

Dubai has a total area of 4291 square kilometres (km^2). Of this total area, however, only about 15% was urbanized in the year 2011 area ($672km^2$) (Roads and Transport Authority, 2010a).[11] Geographically, Dubai has been divided into 221 communities, with a community in the urban area having an average size of 5 km^2. Socio-demographic characteristics within a community are

[11] There are many definitions of an urban area. Here it is assumed that an urban community has a population and employment density of more than 500 persons per square kilometre.

also similar, but the demographic distribution varies across communities. These boundaries are also used for administrative purposes, and to publish yearly demographic and socio-economic characteristics issued by Dubai Statistics Center, Dubai Municipality and other government organizations. In this study, the term neighbourhood and community are used interchangeably. Figure 4.2 illustrates the land use distribution in Dubai, showing that most areas are mixed land-use developments.

The population density in Dubai is presented in

Figure 4-3. The figure indicates that the most densely occupied communities in Dubai are located to the north-east, the north and the west side of the emirate, but that the majority of neighbourhoods in Dubai are sparsely populated. The figure shows that the population in the emirate is concentrated in a spatially narrow strip compared to the total area of Dubai. Figure 4-4 represents this numerically, indicating that around 30% of the communities are not occupied, accounting for 57% of the total area of Dubai. In addition, 43% of the communities (accounting for 85% of Dubai's area) have a population density of less than 100 persons per square kilometre.

Figure 4-2: The existing land use in Dubai (source: Dubai Municipality, 2010)

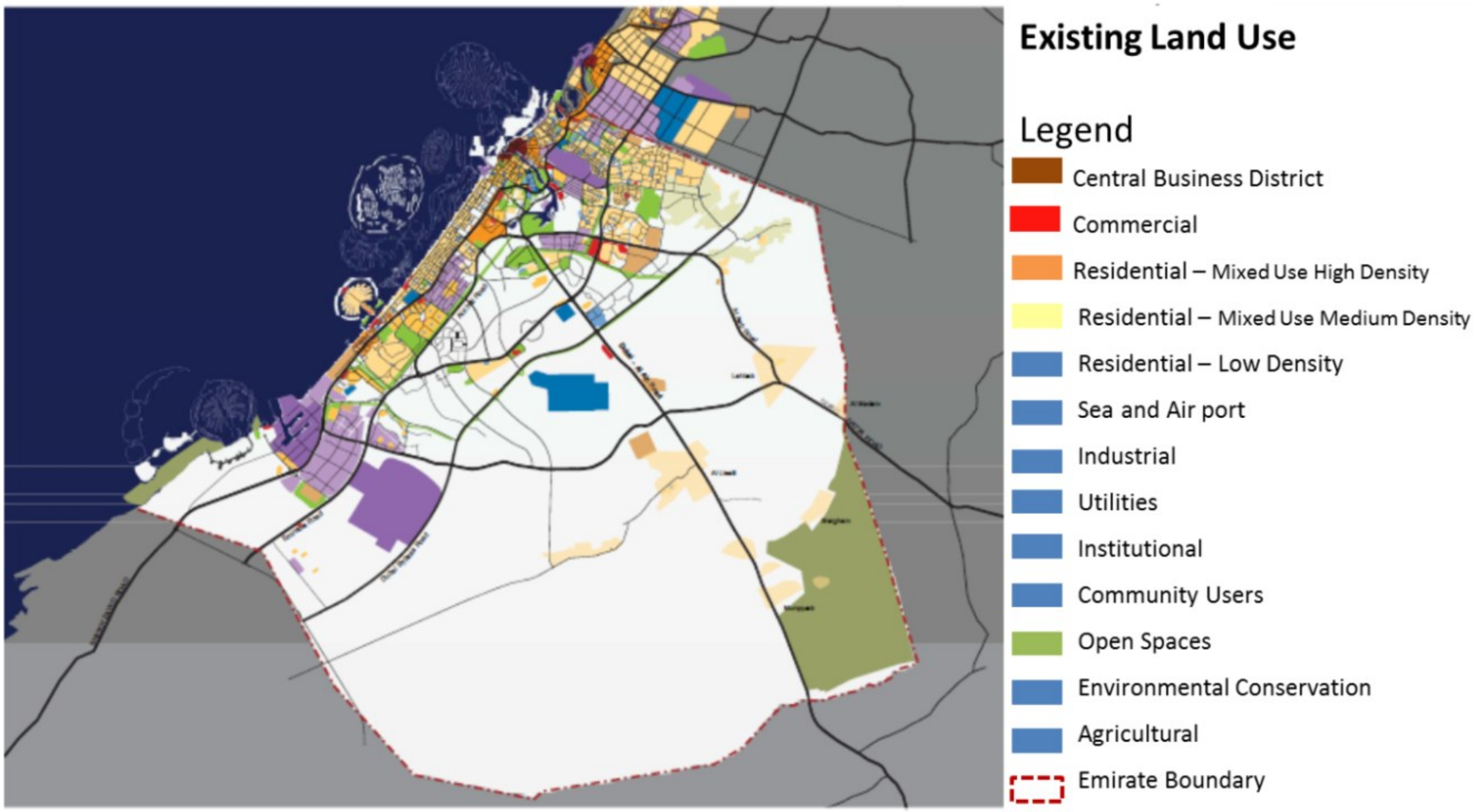


Figure 4-3: Population density in Dubai as of 2011 (source: self-produced graph based on data provided by Dubai Statistics Center, 2012)

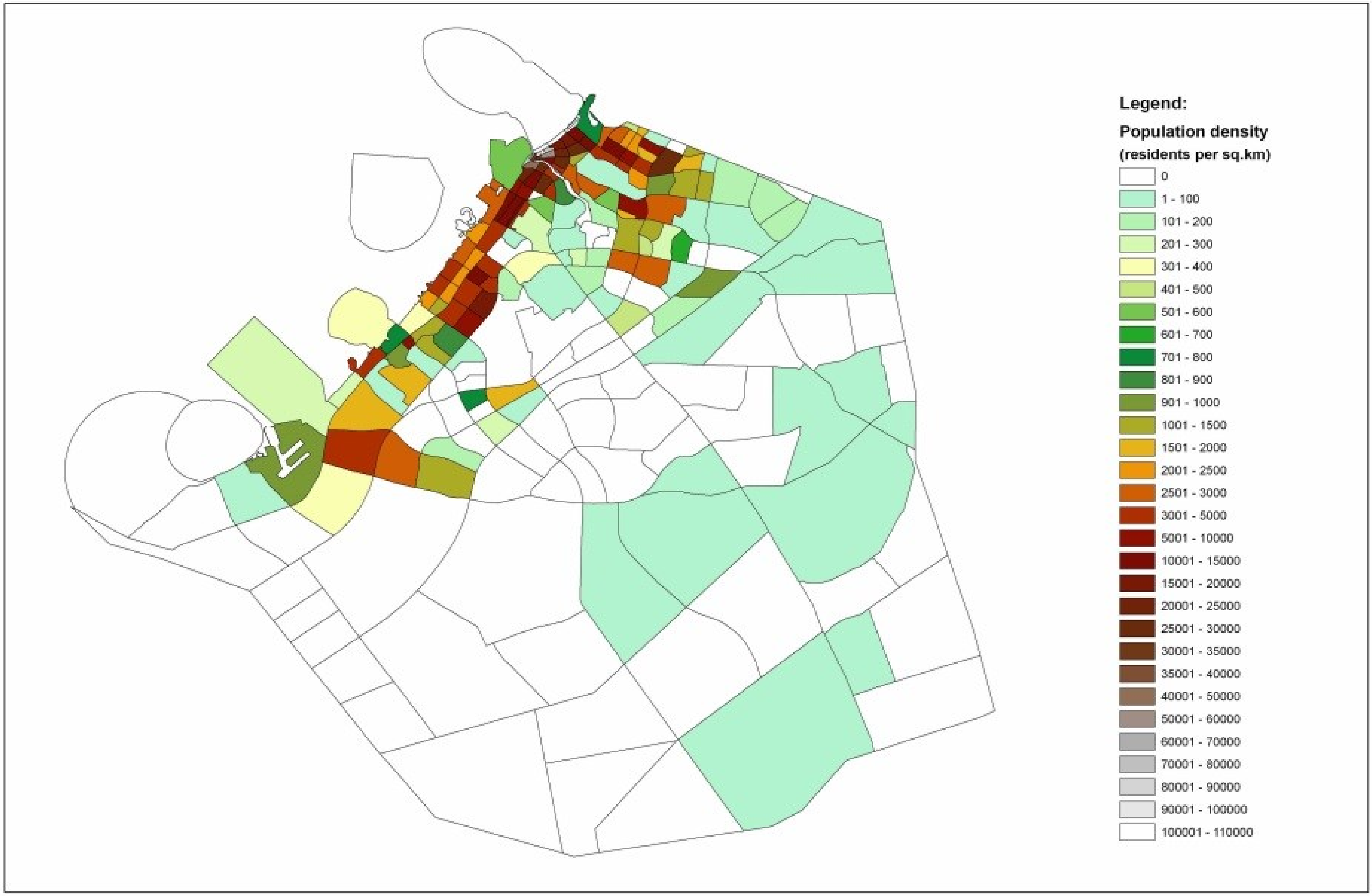


Figure 4-4: The distribution of population density in terms of land area in the year 2011 (source: self-produced graph based on data provided by Dubai Statistics Center, 2012)

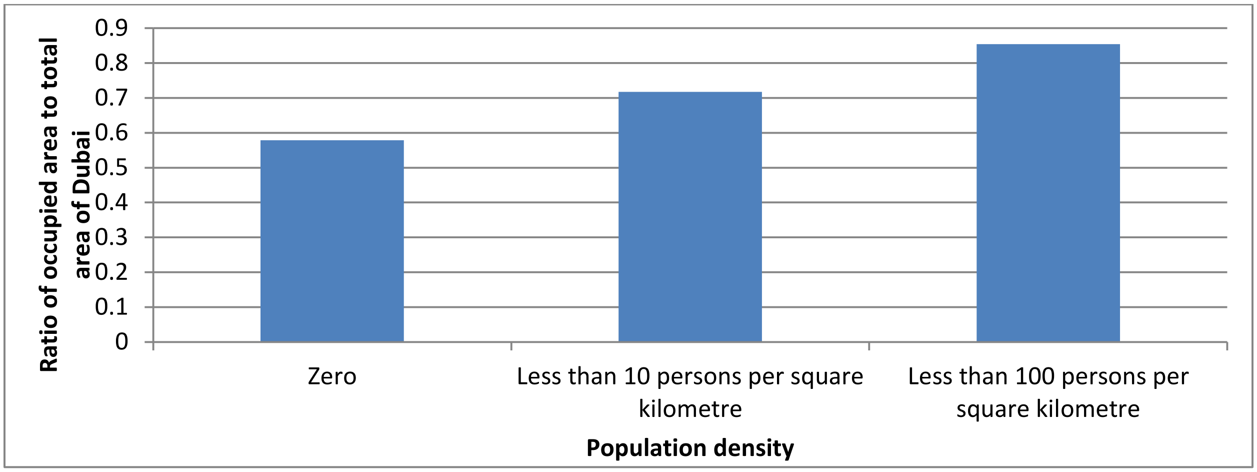


Figure 4-5 illustrates the employment density in Dubai in the year 2011. The old CBD area has the largest employment density, followed by the newly developed communities and other industrial zones. Figure 4-6 reveals that 65% of the communities in Dubai (89% of total area) do not host any employment and that 69% of communities (95% of total area) contain less than 100 employees per square kilometre. The distribution of the employment density is therefore even more compact than the population distribution and indicates that there is a very strong spatial concentration of activities.

Figure 4-7 also shows that the main shopping malls in Dubai (each consisting of around 200 shops to over 1,200 shops) are situated within almost a single corridor.

Figure 4-5: Employment density in Dubai in the year 2011 (source: self-produced graph based on data provided by Dubai Statistics Center, 2011b)

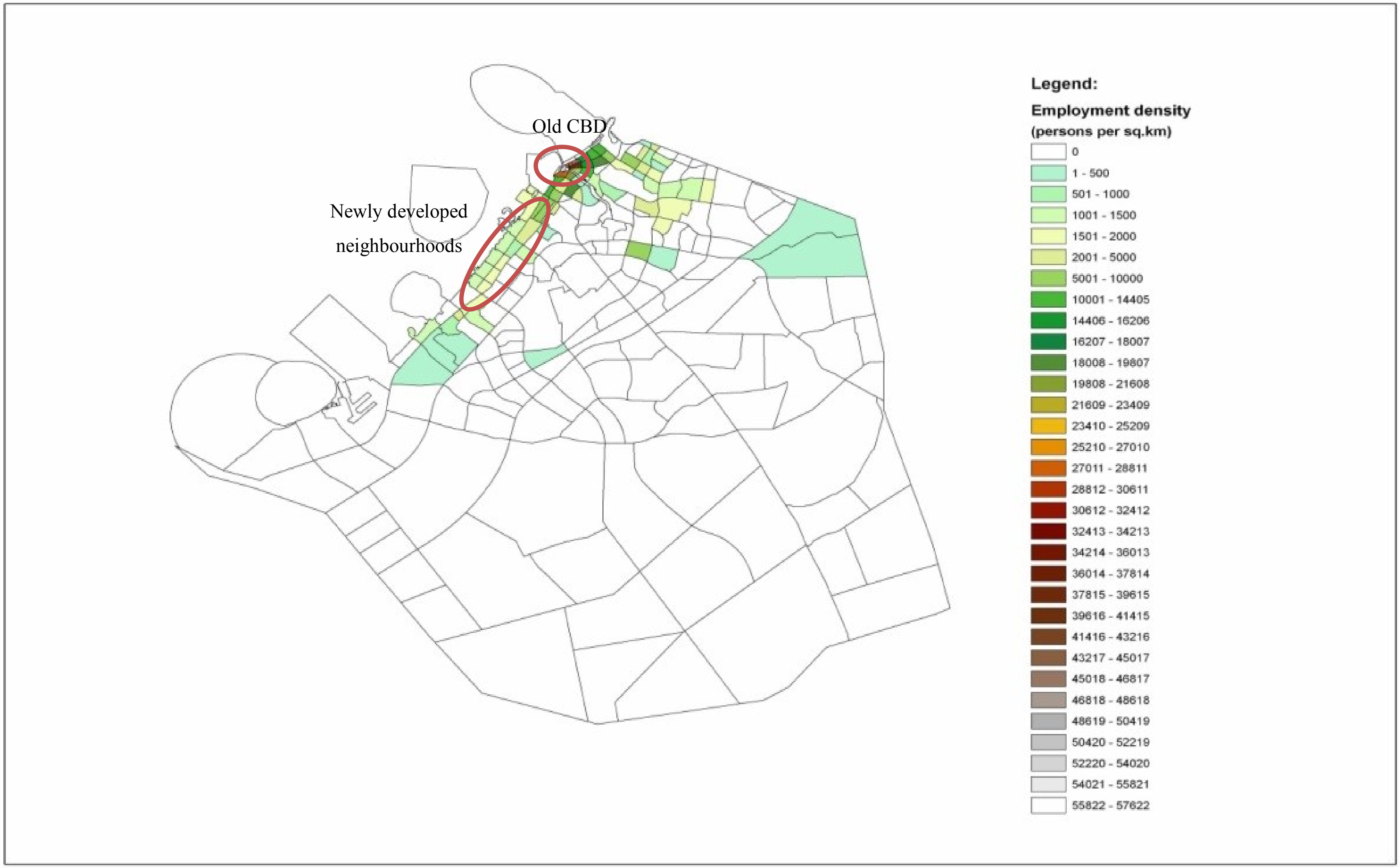


Figure 4-6: The distribution of employment density in terms of land area in the year 2011 (source: self-produced graph based on data provided by Dubai Statistics Center, 2011b)

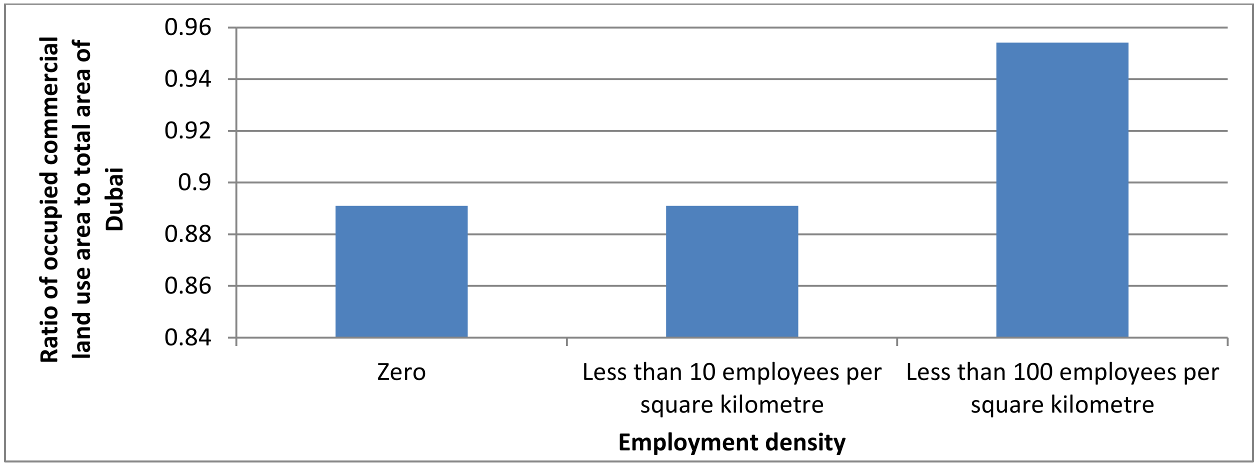
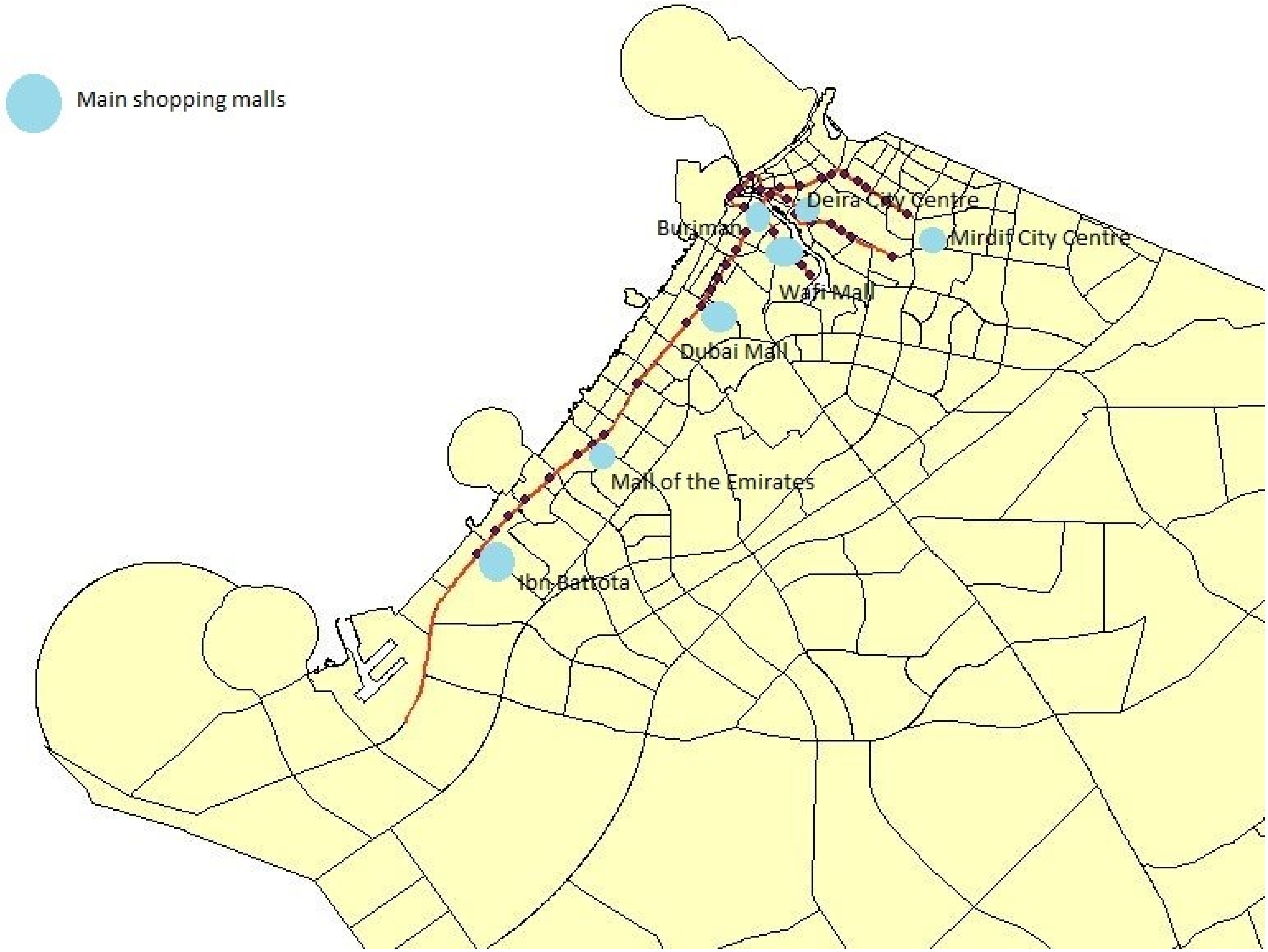


Figure 4-7: The distribution of the main shopping malls in Dubai (source: Dubai Municipality, 2010)



4.3 Transport supply and demand

The high growth in Dubai since the year 2000 has led to dramatic increases in mobility requirements. As a result, Dubai established the Roads and Transport Authority in November 2005 and transferred to it the transport responsibilities from relevant departments within Dubai Municipality and Dubai Police. RTA is responsible for providing transport facilities (such as roads, railways and buses) both within Dubai and to neighbouring emirates.

RTA consists of three central sectors and four operational agencies, linked by a service level agreement.[12] While the central sectors provide the strategic transport plans in Dubai, decide on the level of integration between different modes, and provide technical and administrative support to the operational agencies, the latter implement the projects and provide services. For example, while public buses are planned by RTA, the operations are outsourced. A set of key performance indicators has been developed to monitor the level of service provided to commuters.

To cater for the high travel demands in a car-oriented city – car ownership reached 581 per 1000 population in 2010 (Dubai Statistics Center, 2010) – the emirate has been developing the public transport network to shift a proportion of car trips to public transport modes. While buses and marine transport are being continuously enhanced, Dubai announced the construction of the first metro system in the Middle East in the year 2005, with the metro opening in September 2009. The supply of the transport infrastructure in Dubai is illustrated in

Figure 4-8, while the demand for private transport and for public buses and the metro is presented in

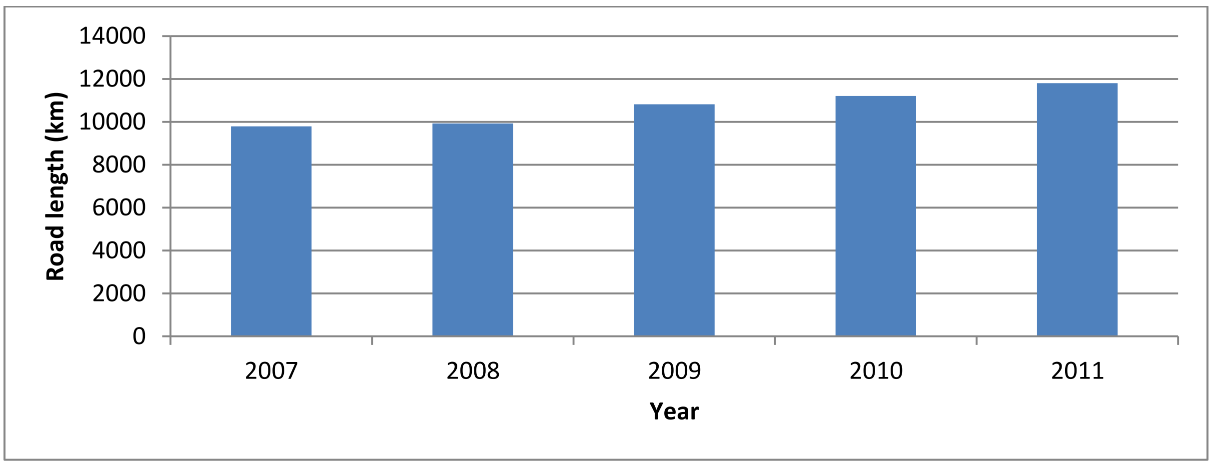
[12] The sectors are Strategy and Corporate Governance, Corporate Technical Support Services and Corporate Administrative Support Services, whereas the operational agencies are the Traffic and Roads Agency, Rail Agency, Public Transport Agency and Licensing Agency.

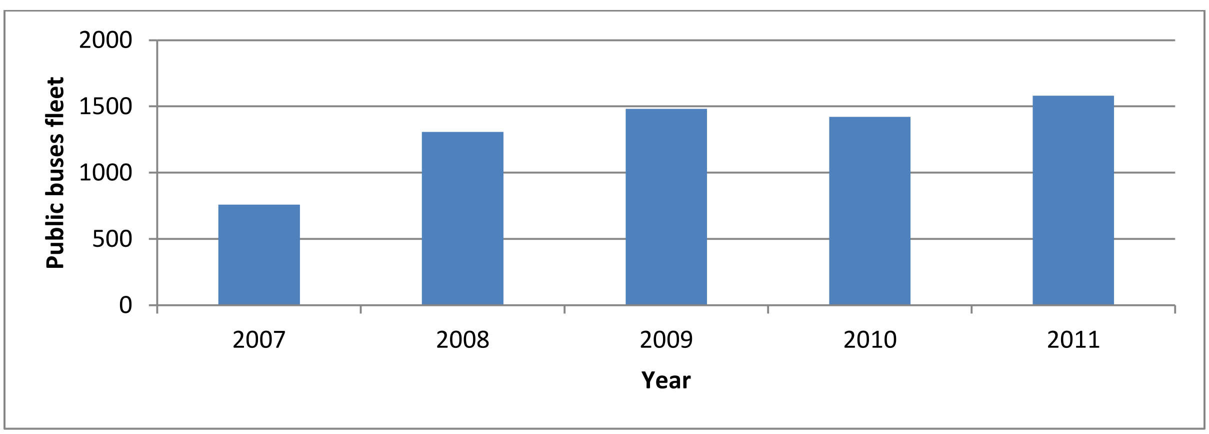
Figure 4-9 and Figure 4-10, respectively. It can be observed that there have been considerable increases in road length and the supply of buses, and also that the metro started operations with just a few stations but subsequently opened more stations the following year.

Due to the growth in the public transport network, the modal share of public transport to private transport trips increased from 6% in the year 2006 to about 10% in the year 2011 (Figure 4-11). It can also be observed that while the increase in the public transport trips is normal until the year 2009, the opening of new stations on the metro led to a larger shift from private to public transport services in the year 2010. This is not surprising, since the metro has introduced a new and more convenient transport service compared to the public buses.

Nevertheless, the modal share of public transport trips in Dubai is still low. There are four main factors that explain this. First, the public transport network is not yet spread widely, which leads to low accessibility and connectivity from origin to destination and low user preference compared to private cars. Second, car ownership in Dubai is inexpensive and, since GDP in the emirate is high, dependence on public transport services is reduced. To achieve a significant change in the modal share would therefore require not only enhancements in the network coverage of public transport but also supporting transport policies (such as higher costs for owning and using a car compared to the costs of using public transport). Third, prior to the introduction of the metro, the only two public transport services (buses and a limited amount of marine transport) were considered as low-profile services. Fourth, since commuters from neighbouring emirates are responsible for a large number of trips in Dubai, and given that the connectivity of public transport services between emirates is poor, there is a particular large dependence on private cars among this group of commuters.

Figure 4-8: The growth in the supply of roads, public buses and the metro from the year 2007 to the year 2011 (source: self-produced graphs based on data provided by Roads and Transport Authority, 2012b)





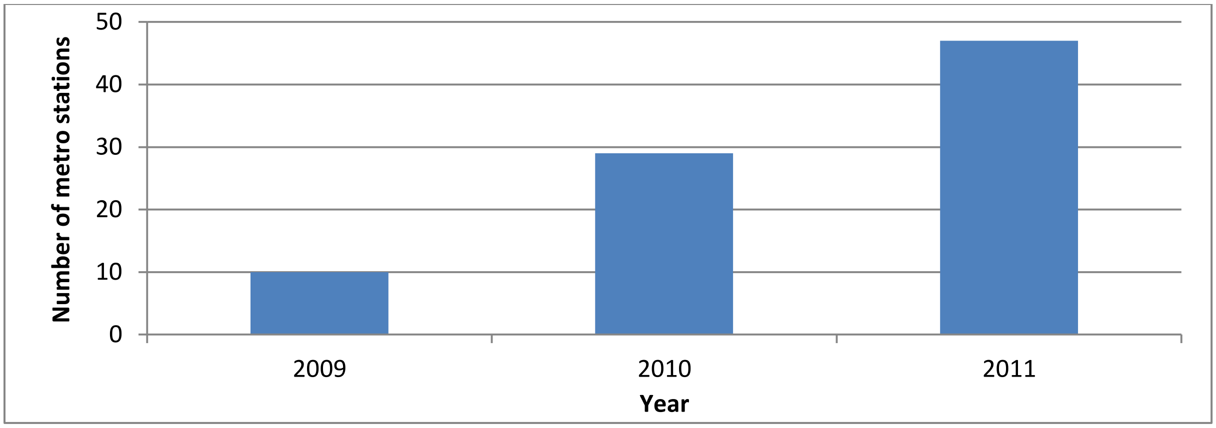


Figure 4-9: Vehicle kilometres travelled by private cars from the year 2007 to the year 2011 (source: Roads and Transport Authority, 2012b)

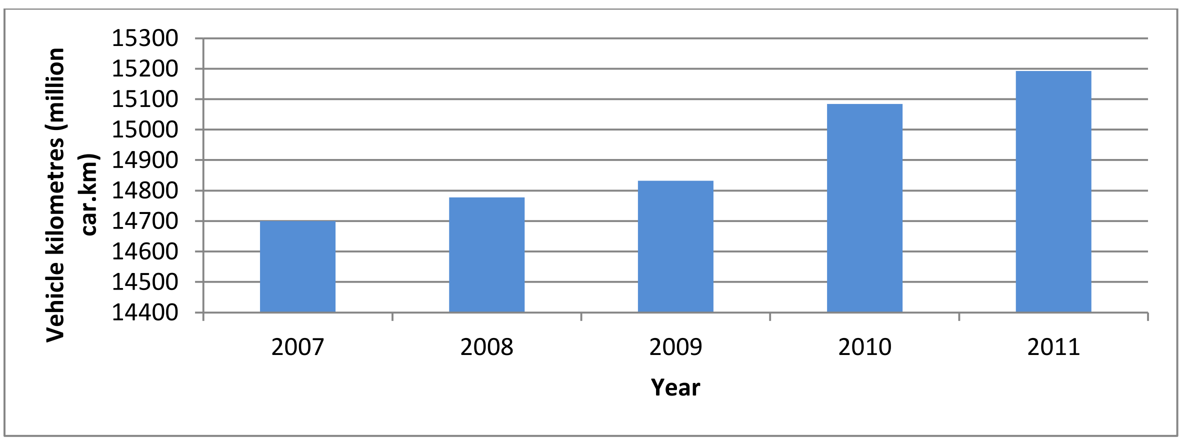


Figure 4-10: Number of passengers using public buses and the metro from the year 2007 to the year 2011 (source: Roads and Transport Authority, 2012b)

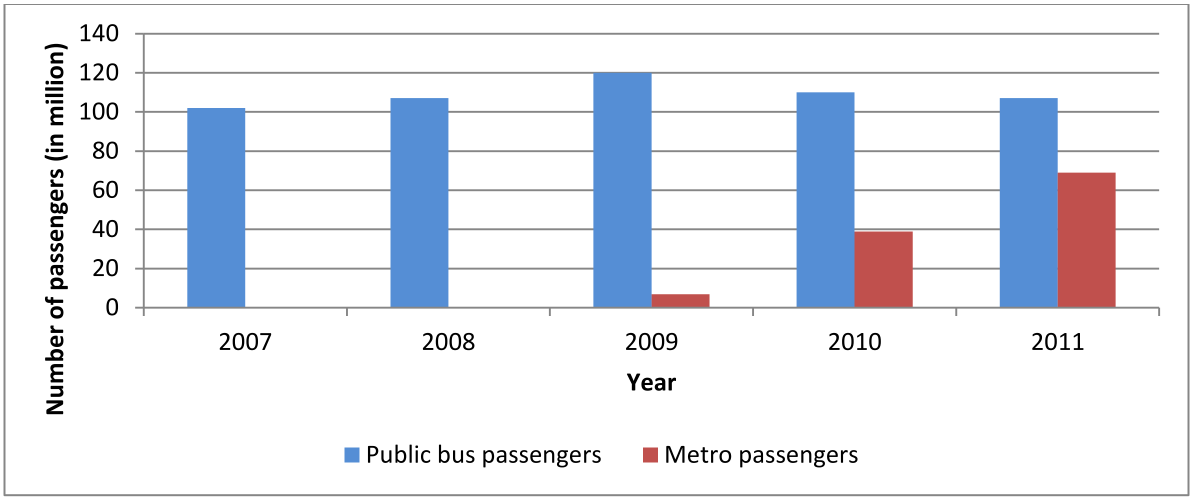
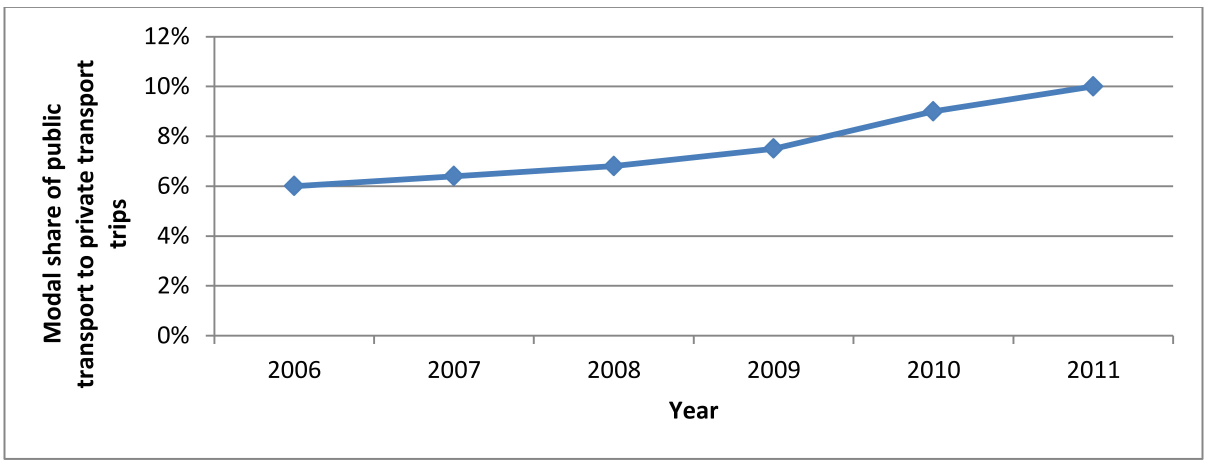


Figure 4-11: The change in the modal share of public transport to private transport trips in Dubai from the year 2006 to the year 2011 (source: self-produced graph based on data provided by Roads and Transport Authority, 2012b)



4.4 Transport innovation – Dubai Metro

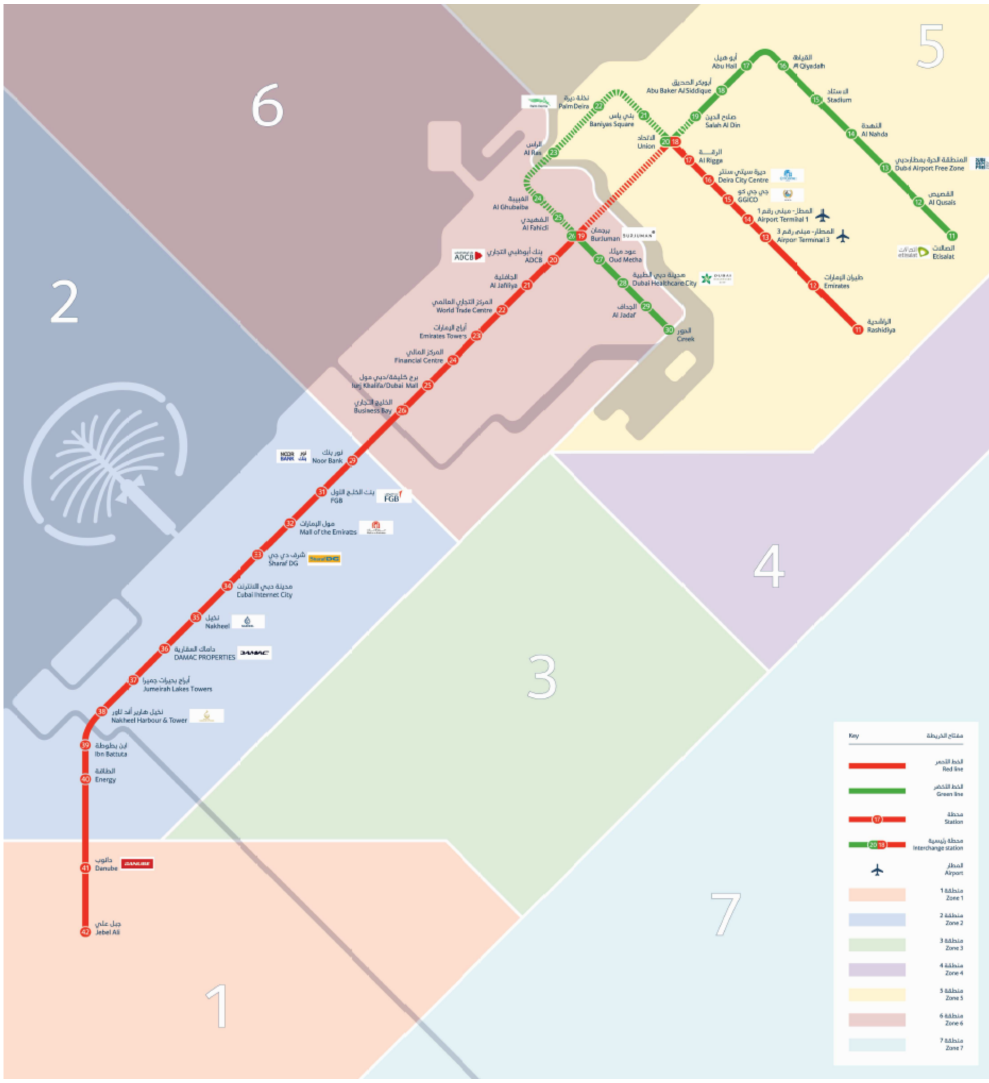
Dubai introduced the first metro in the Middle East in September 2009 by building the Dubai Metro Red line, which was followed in September 2011 by a second line, the Green Line. The metro routes are shown in

Figure 4-12 and the main milestones for the metro project are listed in Table 4-1. The metro was first announced to the public in July 2005, with construction beginning in March 2006. Phase 1 included the opening of ten stations on the Red Line on the 9th September 2009. The remaining 19 stations on the Red Line opened during the year 2010, while the Green Line started operating in September 2011.

Table 4-1: Main milestones for Dubai Metro (source: Dubai Municipality, 2003; and Roads and Transport Authority, 2012a)

Date	Milestone
1992	First study addressing the concept of a metro system in Dubai (not published)
2003	First study to define the route alignment of Dubai Metro
July 2005	First public announcement of the metro
March 2006	Start of the construction work
September 2009	Opening of ten stations on the metro Red Line
Since May 2010	Opening of the rest of stations on the Red Line
September 2011	Start of the operations of the metro Green Line

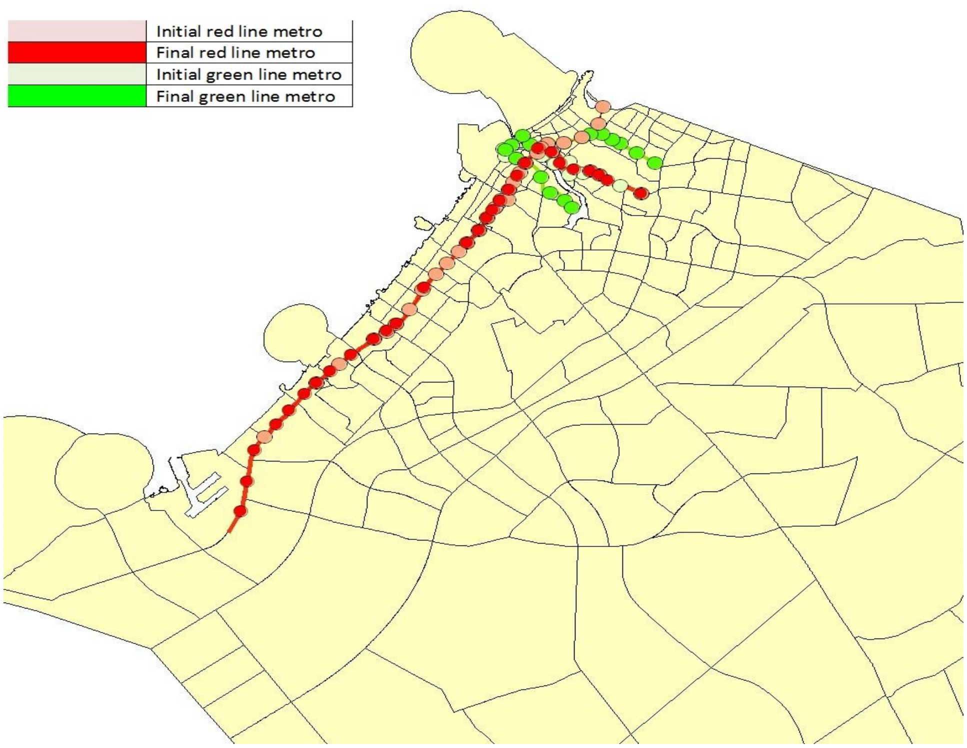
Figure 4-12: A map of the Dubai Metro (source: map provided by the Roads and Transport Authority)



While the idea of implementing a metro in Dubai was initially proposed in a study conducted by Dubai Municipality in 1992, it was not until 2003 that the metro route was planned (Dubai Municipality, 2005). The metro was proposed to cater for the increasing trip demands to the major existing and future employment centres, shopping malls and to serve Dubai Airport. The alignment was chosen to serve the areas with the highest estimated and forecasted trips in Dubai and the route was selected on corridors with available right of way.

Figure 4-13 shows the initial and the final Red and Green metro routes and the station locations. Although most of the constructed route is not different from the initially planned route, some changes have been made. First, the route in the eastern part of Dubai has been diverted south-east where there is a higher employment value, and the previous Red Line is connected to the current Green Line instead. Second, new stations were added to the Green Line to serve the old CBD area in the eastern part of Dubai Creek. Third, the Green Line was extended slightly to the south at the western part of the creek were a new employment centre was announced. Fourth, the initial Green Line route serving the east-south of Dubai Creek was instead connected to the Red Line. The final Red Line route, therefore, is L-shaped whereas the final Green Line is U-shaped. Both the red and the Green Lines integrate at two stations, one on each side of the creek.

Figure 4-13: Initial and final routes of the Dubai Metro Red and Green Lines (source: Dubai Municipality, 2003)



The catchment area around the 29 stations on the final metro Red Line (52 km) consists of old and new employment centres, newly developed mixed use land, at least four of the major shopping malls as well as two of the busiest terminals of the Dubai International Airport. On the other hand, the majority of the 18 stations on the Green Line (22 km) are located in the old CBD area along both sides of Dubai Creek. In 2011, the share of total travel held by the Dubai Metro was 4% and it carried 36% of the public transport passengers (Dubai Statistics Center, 2011a)

4.5 Conclusions

Dubai is a unique city in terms of its rate of development as well as its population structure. The emirate has experienced dramatic growth, and a number of large scale developments have started, although many of these are yet to be completed. In the year 2011, the urban area in Dubai was about 15% of the total area of the emirate. In fact, there is a strong spatial concentration of population and employment distribution in the emirate. In addition, the Dubai Metro serves areas with the most commercial activities (employment and shopping commuters). Given the above, it was decided to check for differences between the treated and control communities and it was found that differences in the employment and shopping distribution in the datasets used in the study need to be controlled for in order to avoid bias.

The diverse mix of nationalities in Dubai implies that the cultural differences and preferences towards transport modes can be quite different. This in turn may lead to varying impacts of the metro on property values located in different communities. Although estimating the effect of the metro on neighbourhoods populated by different demographic groups would add value to this study, as will be shown in the following chapters, the obtained property data is spatially restricted which in turn limits the difference across population groups. Nevertheless, an attempt will be made to include the travel preferences of a number of population groups by measuring the impact on property values of the change in the GC of travel due to the introduction of the metro. Chapter 6 explains how the GC value is calculated and this issue is considered.

With regards to the characteristics of the transport infrastructure in Dubai, it was noted that the supply of public transport has increased dramatically in just a few years, as has the supply of roads. Despite all the measures taken to shift demand to other modes, Dubai remains an overwhelmingly car-dependent city, partly for cultural reasons and partly because around a quarter of the trips in Dubai during peak hours are from neighbouring emirates, to which there are poor public transport links. This being the case, it would be expected that the metro would have only a limited impact on the value of properties, as has been the case in other car oriented cities (see the discussion in chapter 3).

In summary, the results of this study need to consider the context area of Dubai as well as the specific characteristics of the modal share in the emirate. While the next chapter discusses the design rules that were established for the study, as well as the selected empirical methods, the subsequent chapter presents the obtained and calculated transport and property data and selects the suitable datasets for analysis in the context of the defined study dimensions.

Chapter 5. STUDY DESIGN RULES AND THE SELECTED EMPIRICAL METHODS

Following the review of the design context in studies estimating the effect of railways on property values (chapter 2) and the introduction to the emirate of Dubai (chapter 4), this chapter sets out the study dimensions for this work. The potential for estimation bias is reduced by defining the generic boundaries for the property data that will be used to select the most suitable datasets for analysis, as well as to select suitable methodologies. In particular, the chapter covers the data arrangement, time period of property data observations, data structure and empirical methods.

While section 5.1 explains the data arrangement and defines the properties of the treated and control groups, section 5.2 defines the study time period, both before and after the opening of the Dubai Metro. As is evident from the literature, the datasets can be structured in different ways. This study considers repeated cross-sectional data and discusses the construction of pseudo panel data in section 5.3. Section 5.4 sets out and justifies the empirical methods chosen for this study. Finally, section 5.5 concludes the chapter.

5.1 Definition of treated and control properties

Since this study attempts to examine the effect of the Dubai Metro on the value of properties, we classify properties that are close to a metro station (i.e. within a particular catchment area), or which are affected by it, are classified as treated and those that are not affected are classified as control. This classification is required for the use of one of the advanced regression based methods (difference-in-differences), and has the advantage of making it possible to distinguish the impact on properties located within the influence area of the treatment from those further away, even when employing other empirical methods. This section explains further how treated and control properties are defined in this study.

As discussed in chapter 2, the impact radius of railways on property values varies across geographical regions and property types. The majority of previous empirical work has chosen a catchment area based on the judgement of the researcher, either through knowledge of the case study region, or by defining the catchment that best fits their data. In many cases the catchment area falls between 1 km and 2 km from the railway station (e.g. Agostini and Palmucci, 2008; Billings, 2011; Du and Mulley, 2006; Dubé et al., 2014). For example, Agostini and Palmucci (2008) employ a distance band of no more than 1 km from railway stations, since up to 90% of the demand on the metro comes from within this distance. Billings (2011), on the other hand, defines a maximum radius of 1 mile (1.6 km) around a station, because that produces the most reasonable number of treated and control properties considering the scale of the city. In a similar fashion, Dubé et al. (2013) classify properties to catchment areas that allow them to obtain enough records per area; treated properties are those located within 0 to 0.5 km, more than 0.5 km to 1 km and more than 1 km to 1.5 km.

In studying the impact of a reduction in distance to a railway station on the value of residential properties, Gibbons and Machin (2005) separate the effect on the value of treated properties located within 2 km of a station and those located at a greater distance, compared to the effect on control properties. The 2 km threshold is selected assuming a maximum walking time of 30 minutes to a station as well as using a statistical search (kernel regression) that fits their property data. Others do not specifically explain the choice of their thresholds, such as Bowes and Ihlanfeldt (2001) who use different catchments starting at 0.25 miles (0.4 km) to 3 miles (4.8 km). In summary, the literature suggests that the selection of the catchment area(s) threshold(s) around a railway varies across case studies and depends, in many cases, on the available property data.

Since the impact radius of the Dubai Metro has not been identified in previous studies, we used the passenger transfer behaviour to a station to select a reasonable distance band around the station. RTA has been conducting surveys of Dubai Metro users since the year 2010. The 2010 survey suggests that, on average, 53% of passengers walk to stations (Figure 5-1), of which 94%

walk 15 minutes or less (Figure 5-2). Assuming a walking speed of between 4 kilometres per hour (km/hr) to 5 km/hr, as specified in the Dubai Strategic Transportation Model (DSTM) of the RTA, a 15-minute walking time translates to 1 km and 1.25 km distances, respectively.

Figure 5-1: Transfer modes to metro stations in the year 2010 (source: Roads and Transport Authority, 2010b; Roads and Transport Authority, 2011)

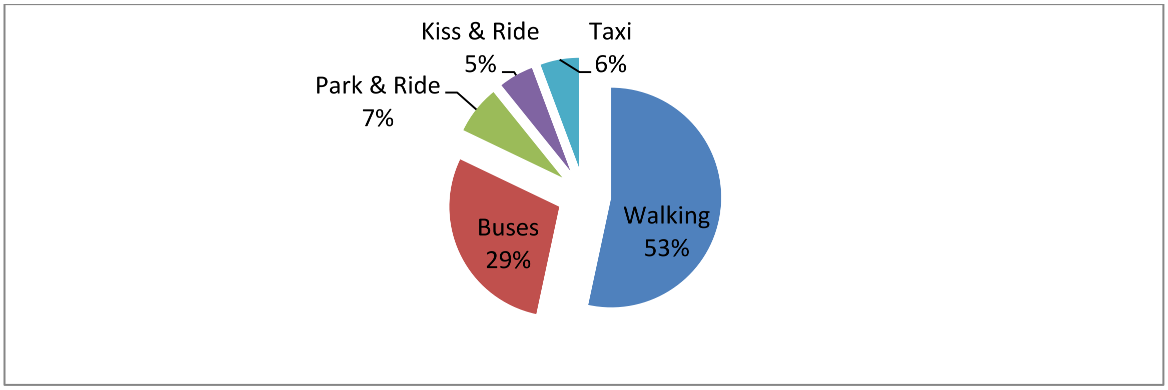
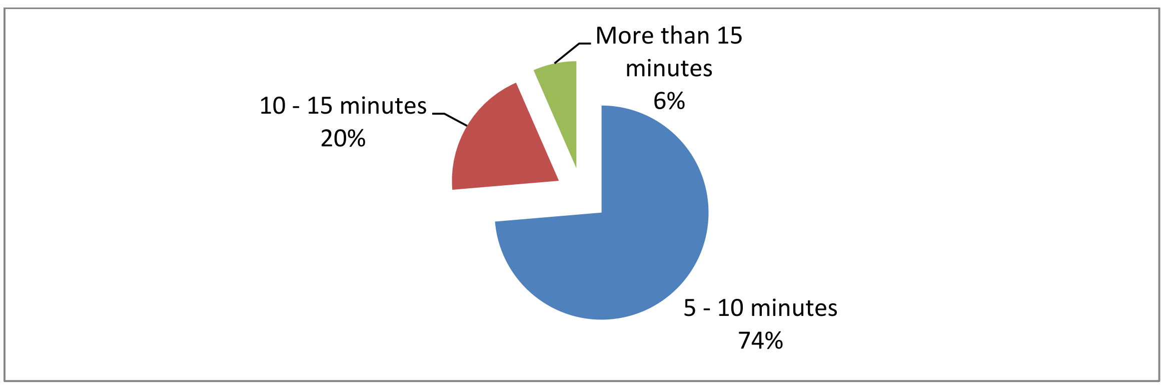


Figure 5-2: Distribution of metro users' walking times in the year 2010 (source: Roads and Transport Authority, 2010b; Roads and Transport Authority, 2011)



In addition, around 47% of users commute to metro stations using motorized modes; i.e. buses, taxis or cars (using kiss-and-ride and park-and-ride facilities). RTA also indicates that the majority (around 80%) of metro users connect to the metro within the community in which it serves (i.e. within an average radius of 1.5 km). Based on this information, a maximum threshold of 1.5 km has been set as the catchment area around the metro stations in Dubai. Properties within these catchment areas were also checked to ensure their accessibility through at least one of the following connections: walking paths, pedestrian bridges, roads or feeder buses.

Ideally, to compare the effect of a treatment on a given economic activity, one tries to compare areas that are similar in all aspects except for the exposure to the treatment; these areas are referred to as 'control' areas. If there are significant differences between treated and control areas (as is the case in the datasets in this study), the term "control" may not necessarily reflect the fact the fact that the absence of treatment is not the only difference between the control areas and the treated areas. Lane et al. (2004) refer to these areas as 'reference' areas.

It can be argued that the definition of 'control' or 'reference' area is related to the type of methodology used. For example, the differences between treated and control or reference groups are not relevant in DID models, given that in the absence of the treatment, the change in the economic activity remains parallel in the treated and control or reference groups as well as in the same groups before and after the treatment. In this situation, any bias due to the differences between the two groups is eliminated. In such cases, In such cases, therefore, it remains appropriate to use the term 'control group' since the areas that are not exposed to the treatment are controlled of the treatment effect. In HP models, however, if there are differences in the conditions of the treated and untreated areas, the term 'reference group' is more appropriate and, in addition, relevant factors need to be controlled for in the estimation models so as to avoid bias. Since this study employs the DID methodology, therefore, the term 'control group' is used.

It is worth mentioning that the choice of the control group differs across studies. Most researchers define the control group as all the properties that are not affected by the transport service (e.g. Agostini and Palmucci, 2008; Ahlfeldt, 2013; Gibbons and Machin, 2005; McDonald and Osuji, 1995). On the other hand, a limited number of researchers who were able to obtain enough property observations across the study area, have used a propensity score matching (PSM) method to define control properties (Billings, 2011; Concas, 2012).

For example, Billings (2011) defines the control groups in three different ways: the first control groups are the areas that were initially selected to be served by a light railway (LRT) station but were later eliminated, the second are all other control areas in the city and the third are selected based on a PSM method. In the PSM method, the control groups are those that have similar neighbourhood characteristics (e.g. the population density, the income levels and the demographic distribution) as those for the treated group before the announcement of the LRT. The author finds similar estimates of the effect of proximity to a LRT station using the different definitions of control groups.

Another example is Concas (2012), who employs a DID method to estimate the effect of proximity to the highway on the value of properties, and who depends on the socio-demographic and housing characteristics to compare the treated and control property groups. Property records were available in a large area of the city, therefore the author chose control groups that had similar socio-demographic attributes to the treated groups, and only those properties in the control groups that have similar housing characteristics to the properties in the treated group.

Ideally, propensity score matching would have been used to select control group areas that have similar conditions to the areas exposed to the Dubai Metro. Since the DID method is used in one of the estimation models (please see section 5.4), however, and since the obtained property records have only a limited spatial distribution, all other areas in the dataset that are further than 1.5 km from a station are used as control areas. We also control for the differences between treated and control areas using relevant factors that may affect property values (as listed later in chapter 6).

Similar to the previous empirical work, this study also considers the effect of the metro within smaller catchment areas. As will be discussed in the next chapter, suitable sample sizes of treated and control properties are identified within distance bands[9] of 0.5 km, 1 km and 1.5 km of a metro station and the effect of the metro on the value of properties is tested for each distance band separately. In particular, in one model, all properties located within a 0.5 km radius of a

[9] A straight line distance from a metro station is used for all distance estimates. This is sensible in this case due to the dense road network in Dubai which results in the network distance and the vector distance being very close values (with an average difference of about 11% as advised by RTA). In addition, since walking pathways are provided along the roads as well as between buildings, walking distances are at least the same as network distances.

station are defined as treated and all properties at further distances as control. In the second model, however, all properties located between 0 km and 1 km of a station are considered treated, with all others being control. In the third model, properties located within 1.5 km of a station are treated and properties located at further distances are control.

5.2 Definition of pre- and post-treatment

In a similar way to the need to classify properties as treated or control, data for pre- and post-treatment is required, both for the DID method, and in order to control for the effect of time on property values. The datasets used in this study (see chapter 6) contain data from the year 2007 to the year 2011, hence the property data is obtained during the construction of the metro until after the operations. This short time span means that it is only possible to estimate the short-term effect of the Dubai Metro. Ideally, we would have wanted to obtain data from pre-announcement of the metro until after operations started.

There are three ways to define the treatment time for this study. The first option considers the months before operations start as pre-treatment months (i.e. from 2007 to August 2009), and from the month of opening onwards as post-treatment (i.e. September 2009 to 2011). The first option cuts the treatment effect at the exact time of opening and assumes that the impact occurs exactly at that time. Since the transaction date is not available for the majority of the property observations used, however, this option is not valid in this case.

While the second option considers the start year of the operations (2009) as a post-treatment year, the third option considers the start year of the operations (2009) as a pre-treatment year. The second option assumes that the effect of the metro operations on property values starts in the same year of operations. The third option assumes that there was no significant capitalization effect of the metro on property values in the year in which operations started, but this occurred at some date post-opening. This assumption is the most reasonable in this case for the following reasons:

First, the Red Line was not fully operational in the year 2009 since only 10 stations out of 29 opened during that year, and it happened that all the treated properties obtained in the datasets were located within 1.5 km of stations that operated in the year 2010. In addition, all the control properties in the datasets were located at a greater distance from any stations that were planned or in operation at the time the study was conducted. This is because the property records obtained were located in the newly developed communities in Dubai, where properties experienced the largest number of transactions, where the majority of listings were available and where a large number of property records were registered, while the majority of stations that operated in the year 2009 were located in the older and less active parts of the city. Hence one may consider the year 2009 as being pre-treatment.[15]

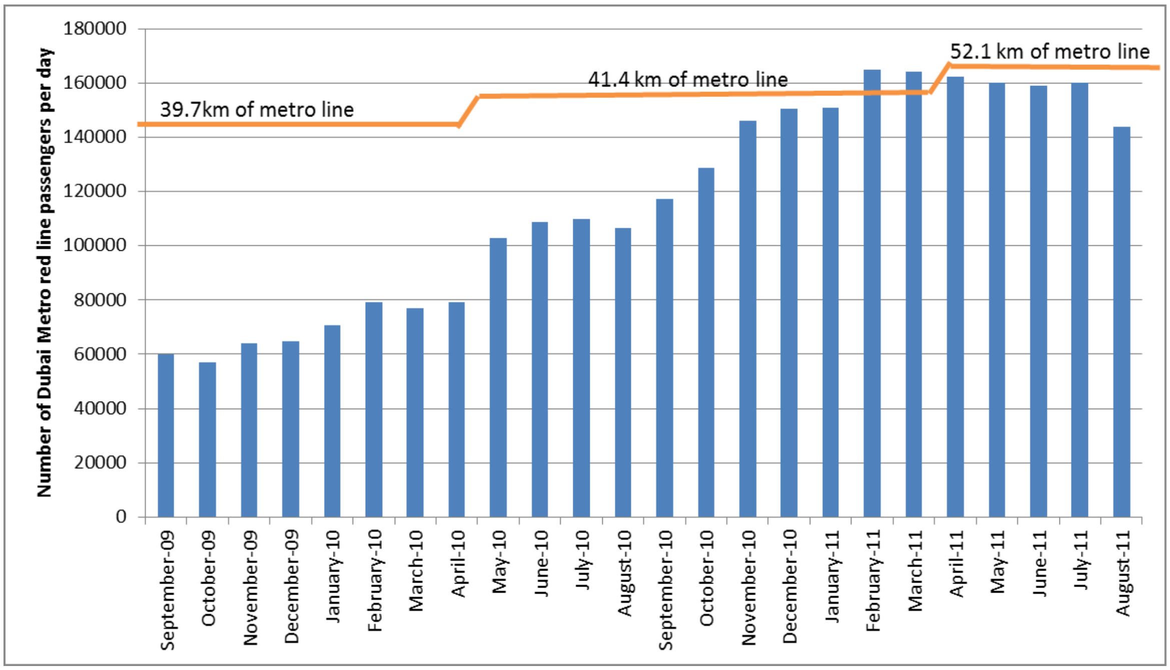
Second, the Dubai Metro is the first of its kind in the Middle East and it operates in a city with a low share of public transport, therefore residents' perceptions of the metro and the anticipated effect on property values in the first three to four months of operations may not accord with those in cities in which residents have experience with and awareness of rail systems and the impact of such systems on property values. In other words, the capitalization effect of the metro on property values may not have been realized before the majority of stations opened and the system started to stabilize.

An indication of this effect can be obtained by examining the daily use of the metro, as in Figure 5-3. Here, an initial jump can be observed, followed by a steady increase in the number of metro users per day after February 2010, although additional stations on the metro Red Line (i.e. additional to the ten opened at the inauguration in September 2009) started operations only after May 2010. This suggests that the metro started stabilizing after around five months of operations. This also supports the choice of the third option and therefore 2009 is designated as a pre-treatment year.

[15] The reader is referred to the next section for the distribution of the obtained property records in our datasets.

In summary, the choice of the pre-treatment cut off year depends on the time of operations, the stabilization period, the behaviour of commuters towards the transport mode as well as the development of market experience related to the effect of the railway on property values.

Figure 5-3: The average number of daily metro users from the start of operations (September 2009) till two years after, together with the total kilometres served by the metro (source: self-produced graph based on data provided by Dubai Statistics Center, 2011b)



5.3 Constructing pseudo panel data

Pseudo panel data is constructed by grouping repeated cross-sectional observations into cohorts so that the mean values of the individual observations become the new individuals in the pseudo panel sample. While the majority of researchers use repeated cross-sectional data to estimate the effect of railways on land and property values, some construct pseudo panel data, while a very limited number obtain panel data. The previous empirical work has only considered one data structure without proper justification for the selection or comparison of results generated using more than one data structure.

As was discussed in chapter 2, pseudo panel data can obviate sources of time-invariant unobserved heterogeneity (Cameron and Trivedi, 2005; Collado, 1997; Deaton, 1985; Verbeek and Nijman, 1992), as can repeated cross-sectional data, by adding location-specific effects (e.g. Agostini and Palmucci, 2008; Koster et al., 2010). The unobserved factors of a cohort in the pseudo panel data can vary over time, however, as the individual observations within a cohort may repeat and each property may hold its specific unobserved characteristics. Amongst others, Deaton (1985) and Tsai et al. (2013) argue that if the cohort size is sufficiently large, the unobserved factors of a cohort can be time-invariant. Verbeek and Nijman (1992) suggest a minimum of 100 observations per cohort to consider pseudo panel data to be genuine panel data. One can also argue, however, that the sufficient cohort size may be more accurately estimated percentage wise; in other words, based on the percentage of the observed cohort size compared to the actual cohort size. As no percentage limit is available in the literature, and since the total number of actual properties per building is not available in Dubai, we are not able to test for this.

Some researchers, however, have chosen either to ignore the minimum size of a cohort or to argue that their estimates are nonetheless reliable. For example, Gibbons and Machin (2005) group on average 2.5 households per cohort, although the actual number of properties per cohort is between 10 – 15 (i.e. cohorts contain 25% to 17% of the actual number of properties). In another study, Weis and Axhausen (2009) compare their cohort sizes with the minimum advised in the literature and conclude that although 50% of the cohorts have a sample size of 25 or less, 85% of the observations are within cohorts with a sample size of 100 or more and therefore they deem that their model estimates are reliable.

There are a number of options to create cohorts from repeated cross-sectional data. Previous studies have grouped observations to either one or a number of time-invariant variables. There is no single measure of correct grouping criteria, but the aims are to group individual records to homogenous cohorts (assuming that observations within a group share almost the same unobserved factors), ensure heterogeneity across groups and retain the most optimal sample size.

Increasing the number of observations within a group reduces the overall sample size and may increase within-group heterogeneity.

The different criteria options considered for this study are listed in Table 5-1. These criteria options are all time-invariant and are listed in a descending level of aggregation. The first option is a highly aggregated method that groups properties according to the type of land use. The second option (grouping the records to the Dubai Municipality identified community where a property is situated) is less aggregated, however this option still groups a very large number of properties. Overall, therefore, in options 1 and 2 properties are quite heterogeneous within a group. In addition, given that only a sample of property data is available for this study, the sample sizes in options 1 and 2 are sharply reduced from the original sample size (a maximum of 15 cohorts are obtained from the thousands of records in the repeated cross-sectional data).

Moving to option 3 (grouping records to the building plot identified either by the parcel ID or the building name, whichever is available), the level of aggregation is much smaller and more reasonable. This grouping criterion adds a spatial dimension to the dataset (i.e. distinguishes the effect of property location) and allows one to control for the effect of unobserved building characteristics. Although option 3 retains a much larger sample size compared to the previous two options, there will still probably be a significant degree of heterogeneity within a cohort, especially if the building consists of a large number of properties. As a result, there may be larger within- than between-group variation.

Table 5-1: Grouping criteria options to construct pseudo panel data from repeated cross-sectional data

Grouping criteria	Description	Pros	Cons
1. Land use type	Observations are grouped to the type of land-use	• Time-invariant for the period of time considered in the study	• Too aggregated since the number of land-uses are limited

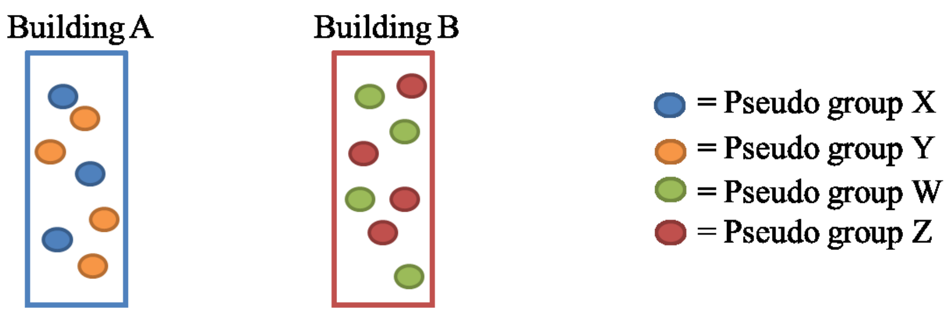
			• Large heterogeneity within a group
2. Community level	Observations are grouped to the communities they are sited in as defined by Dubai Municipality	• Time-invariant • Unobserved factors related to the environment of the community is similar to all records	• Too aggregated since a community is large in size • Limited number of communities • Large heterogeneity within a group
3. Parcel level	Observations are grouped to the parcel number. A parcel is in the size of one building.	• Time-invariant • Reasonable aggregation level • Unobserved characteristics of the building that the property belongs to are common to all flats and unchanged over time. • Retains a reasonable sample size given that data is distributed spatially	• There are too many observations within some parcels and therefore observations are heterogonous in a created group
4. Parcel level and the property size	Observations are grouped to the parcel number as well as to property area or the number of bedrooms.	• Time-invariant • Reasonable aggregation level • Retains a reasonable sample size. • Records within a cohort are expected to be homogenous given that they belong to the same building and share similar property sizes. • Better representation of unobserved factors in each group compared to the other options above.	• When the number of bedrooms is used as a grouping measure, this variable is not included as a covariate in the models, due to endogeneity

To construct more homogenous groups, more than one grouping option can be used. We propose to combine the building in which the property is situated with the property size (option 4). In addition to the advantages of grouping observations to the plot of land, the second criterion based on property size makes it possible to control for unobserved heterogeneity across properties due to their size. The size is defined either by the number of bedrooms, if available, or by a range of area groups. Area groups are classified in sizes of every 150 square metres (sqm) starting with 50 sqm and above; all properties of less than 50 sqm in size are generally studio flats and are hence grouped together. Option 4 tries to balance the trade-off between the sample size and the consistency of the created groups.

In order to finalise the optimal grouping criteria, the variation within- and across-cohorts was tested for options 3 and 4 (appendix B). Although it was found that the within-cohort variation was not always less than the across-cohort variation using grouping option 3, grouping option 4 always resulted in a larger across- than within-cohort variation (i.e. increased homogeneity within the group and increased heterogeneity across them). It was therefore decided to group records by building and property size for all selected datasets.

Figure 5-4 explains the construction of the pseudo panel data in this study. For a given plot of land (a parcel ID or a building), we group properties that share similar sizes (number of bedrooms or range of area sizes) in a pseudo group. For example, records on one-bedroom flats that are situated within building A (represented by the blue circles in the figure) are grouped together in pseudo group X. Similarly, two bed-room flats located in building A (represented by the orange circles), and two bedroom flats located in building B (represented by red circles) are grouped to pseudo groups Y and Z, respectively. A mean value is then calculated for each attribute related to the properties that are located within a given pseudo group, and this mean is then the value of the attribute in the pseudo panel dataset. The pseudo groups repeat over time and hence one record of an attribute is available per pseudo group per year.

Figure 5-4: Illustration of the construction of pseudo panel dataset (source: self-produced graph)



Pseudo group	Year 2007	Year 2008	Year 2009	Year 2010	Year 2011
X	V1	V2	V3	V4	V5
Y	V6	V7	V8	V9	V10
W	V11	V12	V13	V14	V15
Z	V16	V17	V18	V19	V20

To control for the unobserved heterogeneity across properties using the repeated cross-sectional data, the properties were clustered to location and size (similar to option 4). In the absence of a reference to each property in the datasets used for this study, this clustering option is the most disaggregated level of grouping that can be used. Properties were obtained per cluster before and after the operations of the metro and, although the same property may not have gone through a resale or rent after the treatment, other properties with similar characteristics in the same building were observed. In this way, time-invariant unobserved factors were controlled for per cluster. On the other hand, sale and resale of the same cohort is available with time in the pseudo panel data. Although pseudo panel data aims to reduce bias from omitted variables related to the cohort, the data sample size is considerably reduced and therefore this data structure is prone to bias from measurement error.

5.4 Choice of empirical methods

As presented in chapter 2, researchers have used different empirical methods to measure the effect of a transport system on land and property values. The choice of any particular method is closely related to data availability and arrangement. For example, if data before the treatment or for control groups is not available, DID will not be a valid method. For this case study, three methods are potentially considered: these are the HP, GWR, DID and SDID models.

Although the HP and the DID methods can control for unobserved heterogeneity, the former only provides indicative results for the effect of railways on property values because it does not estimate causal relationships. The ability to distinguish the effect of the regressors spatially gives a GWR model an advantage over a HP model in the presence of spatial heterogeneity. GWR, however, is not suitable for this study for reasons related mainly to data availability. First, some of the datasets lack the coordinates of the observations, hence the spatial dimension to the GWR model could not be included, which in turn prevents the application of this method. Second, although records of properties located in hundreds of buildings in Dubai have been obtained, these are clustered spatially in just a few communities.[16] This relatively narrow distribution of property records further limits the use of a spatially varying model, since the model relies heavily on the values of nearby properties to estimate the effect on the dependent variable.

Similarly, SDID requires spatial distribution of property data in order to relate the value of one property to the value of a nearby property and control for spatial differences in values. If in the future data that is more spatially distributed across Dubai is obtained, together with the location ID for all observations, then GWR and SDID could be considered.

Given the above, it is argued that the most suitable empirical methods for this study are the DID and the HP methods. The two considered data structures (repeated cross-sectional and pseudo panel data), combined with the DID and the HP models, are together able to provide consistent average estimates for the effect of the Dubai Metro on property values. It is also possible to compare results across the datasets using the same empirical methods. The detailed structure of each model is presented within the related empirical chapter (chapters 7, 8 and 9).

[16] Reader is referred to the next chapters for details

It is also worth mentioning that in the presence of spatial variation of property data, results from Ordinary Least Square (OLS) methods are biased. To control for this, one may either adopt a feasible Generalized Least Square (GLS) method or use a form of spatial econometric model (e.g. Concas, 2012; Kim and Zhang, 2005; Martinez and Viegas, 2009). In this study, we adopt the first approach across all models.

Most of the existing empirical studies examining the impact of railways on property values have not discussed the appropriate type of estimator for the empirical models (e.g. a random-effects (RE) or a fixed effects (FE) estimator), although the majority have chosen an RE estimator by default. The results using an RE estimator, however, are inconsistent if significant variables are omitted in the regression models. In the presence of omitted variable bias, meanwhile, an FE estimator generates consistent estimates but these may be inefficient if the within-group variation is not large enough. Although conventionally some researchers depend on the Hausman test to decide on the suitable estimator (FE vs. RE), this test is not reliable if the between- and within-group variations are relatively small (Clark and Linzer, 2013; Hahn et al., 2011; Plümper and Troeger, 2007).

A small number of researchers have tested for the consistency of results using various estimator types in different combinations of within- and between-group variations, as well as cohort sizes. For example, Clark and Linzer (2013) use a Monte Carlo Simulation to evaluate the consistency of results, and to decide the most suitable estimator type based on the lowest Root Mean Square Error (RMSE). An aspect of Clark and Linzer's analysis that is applicable to this research is that, in the case of large numbers of small-sized cohorts, low within-variation and low correlation between the dependent variable(s) and the unobserved factors,[17] an RE estimator produces more consistent estimates compared to an FE estimator. Another study, conducted by Tsai et al. (2013), also uses a Monte Carlo Simulation, however on pseudo panel data. They also indicate that, in the case of low to moderate correlation between the unobserved factors and the

[17] The following chapters show that a large number of small-sized cohorts were obtained in this study. In addition, a low level of within-group variations was found. Correlation between unobserved factors and regressors was checked for by conducting a correlated random effects model: the results show low correlations.

regressors, as well as lower within-group variation than between-group variation, the RMSE is lower (and hence the model is more reliable) when using an RE estimator compared to an FE estimator.

Since the between- and within-group variations in the data for this study are low, the Hausman test is not efficient in deciding the most suitable estimator. The recommendations from previous studies that used the results of RMSE to decide on the most suitable estimator type were therefore relied. Given the arrangement of the data in this study in respect to the number of cohorts and between- and within-variations, previous studies suggest that an RE estimator provides more consist estimates than an FE estimator. The findings of this study are, therefore, discussed using an RE estimator, although in appendix F the key results are also reported using an FE estimator.

5.5 Conclusions

As was discussed in chapter 2, a given combination of the spatial and temporal distribution and arrangement of property data is required not only to reduce estimation bias when using conventional empirical methods but also when using some advanced methodologies for estimating the effect of a railway on land or property values. This chapter has set out and justified the study design rules in terms of the metro catchment area, the study time period, the data structure and the suitable empirical methods for estimating the effect of the Dubai Metro on property values. The study dimensions are also checked to confirm that they fit the context area of Dubai as well as the provided data.

By analysing metro users' commute behaviour to a metro station it was possible to define a maximum catchment area comprising a 1.5 km radius around a metro station. This distance band is similar to that used in previous empirical work. In addition, it was observed that the time span considered in other studies examining the effect of a railway on property values varies according to data availability, but that more recent studies tend to obtain data from pre-announcement to post-opening. Due to limitations in the temporal span of the provided property data in Dubai, this study estimates only the short-term effect of the metro (during construction to two-years after operations) and the results are therefore likely to be undervalued. It is argued here that the years up until the start year of the operations are to be considered as being before the treatment occurs. This choice of the treatment time has also been adopted in some previous studies.

The data structure and the empirical methods can have an impact on the potential bias in studies like this one. In the absence of panel data, the literature has shown that using repeated cross-sectional data, corrected for unobserved effects, together with pseudo panel data, may produce consistent estimates of the effect of a railway on land or property values. Both data structures are therefore applied in this study, with their results being compared. The construction of the pseudo panel data is also discussed in this chapter. Finally, among the existing empirical methods used for examining the effect of a transport system, the HP and DID methods are chosen as the most suitable given the data available for this research. While this chapter has presented the study design rules for estimating the effect of the Dubai Metro on property values, the next chapter assesses the obtained land and property datasets and selects the most suitable in light of the study objectives and dimensions. In addition, the next chapter sets out the data related to the transport elements considered for this study.

Chapter 6. TRANSPORT AND PROPERTY DATA

6.1 Introduction

This chapter describes the transport data and the land and property data used in this study. The transport data consists of the variables that measure the accessibility offered by the metro (in distance, travel times and costs) which are obtained either from RTA, or calculated using an ArcGIS system. Records on land and property contain observations on land and property values as well as their location and characteristics. These are obtained from various sources in Dubai, principally the Dubai Real Estate Regulatory Authority (RERA), which is a government entity that holds repeated cross-sectional data on actual sale and rent transactions for different types of land and property. Other sources used for acquiring property data are one of the main real estate consultants (named REIDIN) as well as the Dubai Statistics Center (DSC).

Although there is some limited overlap of land and property data across these data sources, generally, the data sources provide different types of records (land vs. property, transactions vs. listings and repeated cross-sectional data vs. panel data). The advantages and limitations of each dataset are discussed in this chapter in relation to the study design rules presented in the previous chapter. Section 6.2 presents the transport data collected for this study, while Section 6.3 describes and discusses the obtained land and property datasets, the selection of the final datasets, summary statistics in the two data structures as well as the addition of missing attributes. Section 6.4 concludes the chapter with some recommendations for the concerned entities in Dubai regarding land and property data collection and the management of data for future analysis.

6.2 Transport data

Three measures of the accessibility offered by the metro are used to estimate the effect of the metro on property values. Accessibility is here defined in two ways: the physical proximity of a property to a metro station (hence the use of a binary variable for distance and the number of accessible metro stations) and the change in the travel times and costs due to the metro (hence using generalized cost of travel values). These measures are explained in this section.

The first measure used to test for the effect of the metro is a binary value that is equal to one if a property is located within the predefined catchment area around a metro station, and zero otherwise. The second is the number of metro stations that a property has access to within the predefined catchment area. The first two measures are created after calculating the distance from each property to all metro stations. The distance is extracted using an ArcGIS tool on the Dubai network, given that the parcel ID of a property is provided. For cases where no parcel ID is given, distances to nearby metro stations are obtained by matching the building name that a property is located in with the information on the number of stations close to the buildings, as provided by real estate agents and property listing websites.[18] A cross-check of the information provided by different websites was also conducted. Figure 6-1 illustrates the first two measures of accessibility.

The third measure is the change in the generalized cost of travel by private and public transport due to the operations of the metro. The generalized cost of travel (GC) equals the monetary cost of using a transport system (for example the fare to use the metro or the fuel cost to use a car) in addition to the monetary values of in-vehicle and out-of-vehicle travel times (e.g. travel time from origin to a metro station, travel time on the metro, transfer time, if applicable, and time to reach the destination). The GC value is calculated in terms of the costs of travel in UAE currency (AED).

The values of the GC of travel by public transport and private transport are obtained using information from the Dubai Strategic Transport Model (DSTM). Ideally, GC values would have been obtained before the opening of the metro (i.e. pre- and in 2009) and after its opening (i.e.

[18] This information is readily available from the websites of major real estate and property listings. The main websites used to extract extra information were http://www.bayut.com; http://www.bhomes.com/; http://www.propertyfinder.ae; http://www.emaar.com/

post-2009). This information is not available, however, and cannot be back-calculated due to lack of travel time data.

Values of the GC of travel with and without the metro in operation are therefore calculated for the year 2010. This process has the following implications. The comparison of the GC values with and without the metro in the same land use context in Dubai and holding all other parameters constant (e.g. population, employment and shopping distributions, the transport supply of other modes and travel behaviour), has the advantage of revealing the change in the GC solely due to the metro. Nonetheless, one may argue that the reverse effect could occur such that the metro may have affected the distributions of the employment and shopping destinations. Hence keeping external factors constant and measuring the GC values with and without the metro may result in biased estimates. As explained in section 4.4, this argument does not hold in this case.

Before describing how the GC value is calculated, it is worth mentioning that the DSTM captures the different mixture of population groups resident in a community. Table 11-5 in appendix C lists the population groups in DSTM. To account for the differences in the travel preferences and travel behaviour of each population group, each group has a unique transport related parameter obtained through a survey conducted earlier by RTA. It is assumed that the population groups do not change over the time span of the study. It is worth mentioning that the value of time is estimated as the average value of time for all population groups within a community. In this way, we also incorporate the travel preferences of the mixture of population segments within a community, since this can affect the bid value of land and property located at some distance from a metro station (McCann, 2001).

The most disaggregated value of the GC available from the DSTM is per community, calculated from the centroid of the community to all possible destinations connected by private or public transport systems. The equation below, and Figure 6-2, illustrate the concept of estimating the value of the GC of travel. For example, to find the GC of travel using public transport services from a given property, the average travel time from the centroid of the community containing the property to the centroid of all communities in Dubai is first calculated, given that there is a connection between the communities using public transport services. The time is then multiplied by the average value of time for the community to obtain a cost variable of time in AED. Then, the average monetary cost of using the public transport system is added. This provides the average GC of travel from the centre of the community where a property is located to the centre of all possible communities in Dubai. This value is available for the AM and PM peaks and an average of both time periods is calculated to obtain an average peak hour GC value.

In order to find the GC per property, the value of the time needed to travel from the building where a property is situated to the centroid of the community is added to the average GC value. In doing this it is recognised that there is a risk of overestimating the GC of travel values for properties that have a short-cut access to the transport service instead of a connection through the centroid of the community, but since there is no other information available, it is suggested that this approximation is the closest feasible to the real GC value. This is done for the case with and without the metro as well as for using private transport.

$$GC_{i(jc)t} = \frac{\sum_C V_{P.o(c)} \cdot T_{o(c)t} + \sum_C V_{P.in(c)} \cdot T_{in(c)t} + \sum_C V_{P.tr(c)} \cdot T_{tr(c)t} + \sum_C F_{(c)t}}{C} + V_{Pi(jc)} . T_{i(jc)t} \quad (5)$$

where:

$GC_{i(jc)t}$ is the average generalized cost of travel in time t for property i within building j situated in community c to all communities in the emirate (in AED)

$V_{P.o(c)}$ is the average value of out-of-vehicle time across all population groups P to travel from the centroid of community c to another community in the emirate (in AED per hour)

$T_{o(c)t}$ is the average out-of-vehicle travel time in time t from the centroid of community c to another community in the emirate (in hours)

$V_{P.in(c)}$ is the average value of in-vehicle time across all population groups P to travel from the centroid of community c to another community in the emirate (in AED per hour)

$T_{in(c)t}$ is the average in-vehicle travel time in time t from the centroid of community c to another community in the emirate (in hours)

$V_{P.tr(c)}$ is the average value of transfer time between modes across all population groups P to travel from the centroid of community c to another community in the emirate (in AED per hour)

$T_{tr(c)t}$ is the average transfer time in time t between modes from the centroid of community C to another community in the emirate (in hours)

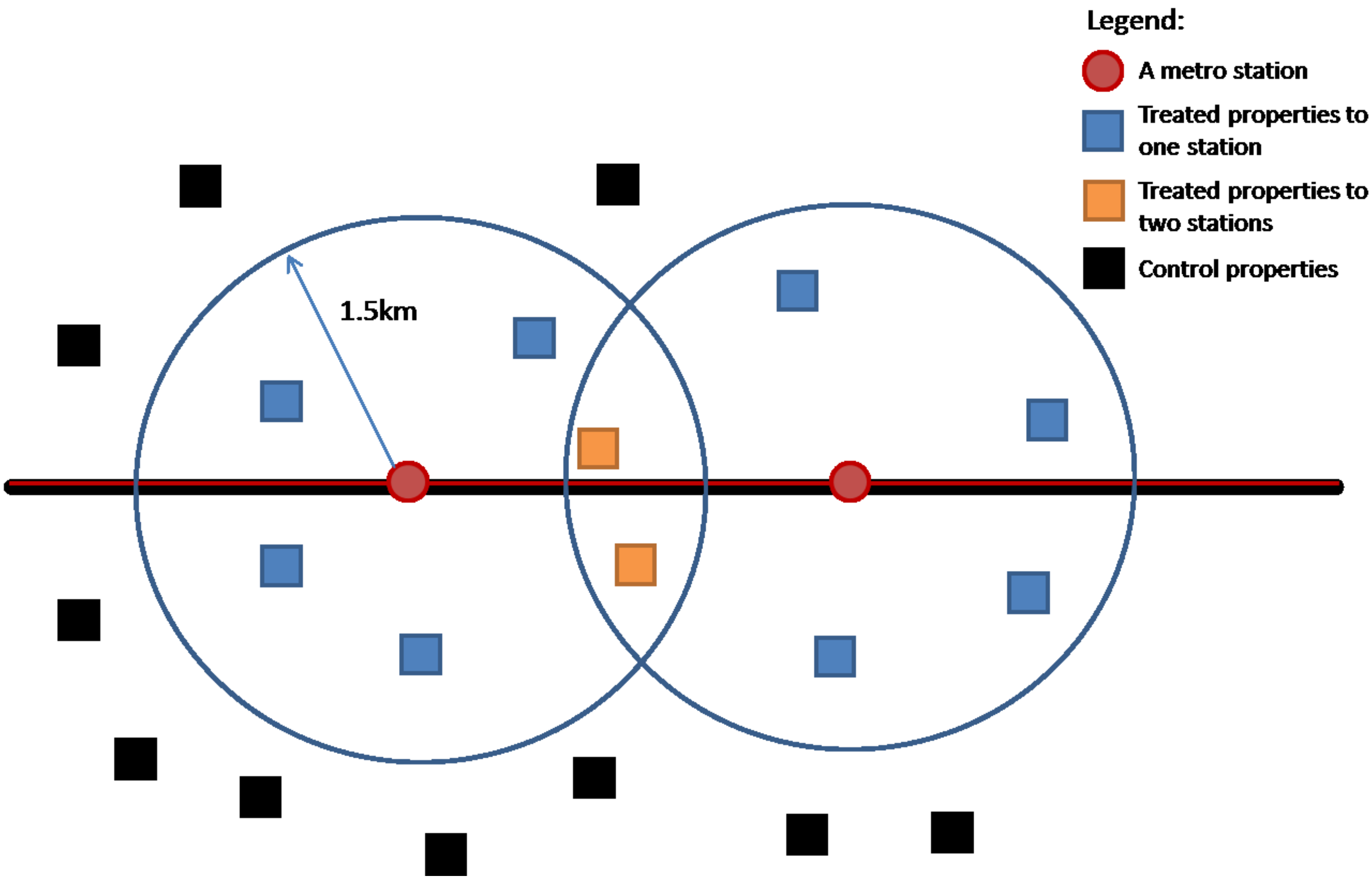
$F_{(c)t}$ is the cost or the fare of using the transport service in time t from the centroid of community c to another community in the emirate (in AED)

$V_{Pi(jc)}$ is the value of the transfer time from property i within building j situated in community c across all population groups P to the centroid of community c (in AED per hour)

$T_{i(jc)t}$ is the transfer time from property i within building j situated in community c to the centroid of community c (in hours)

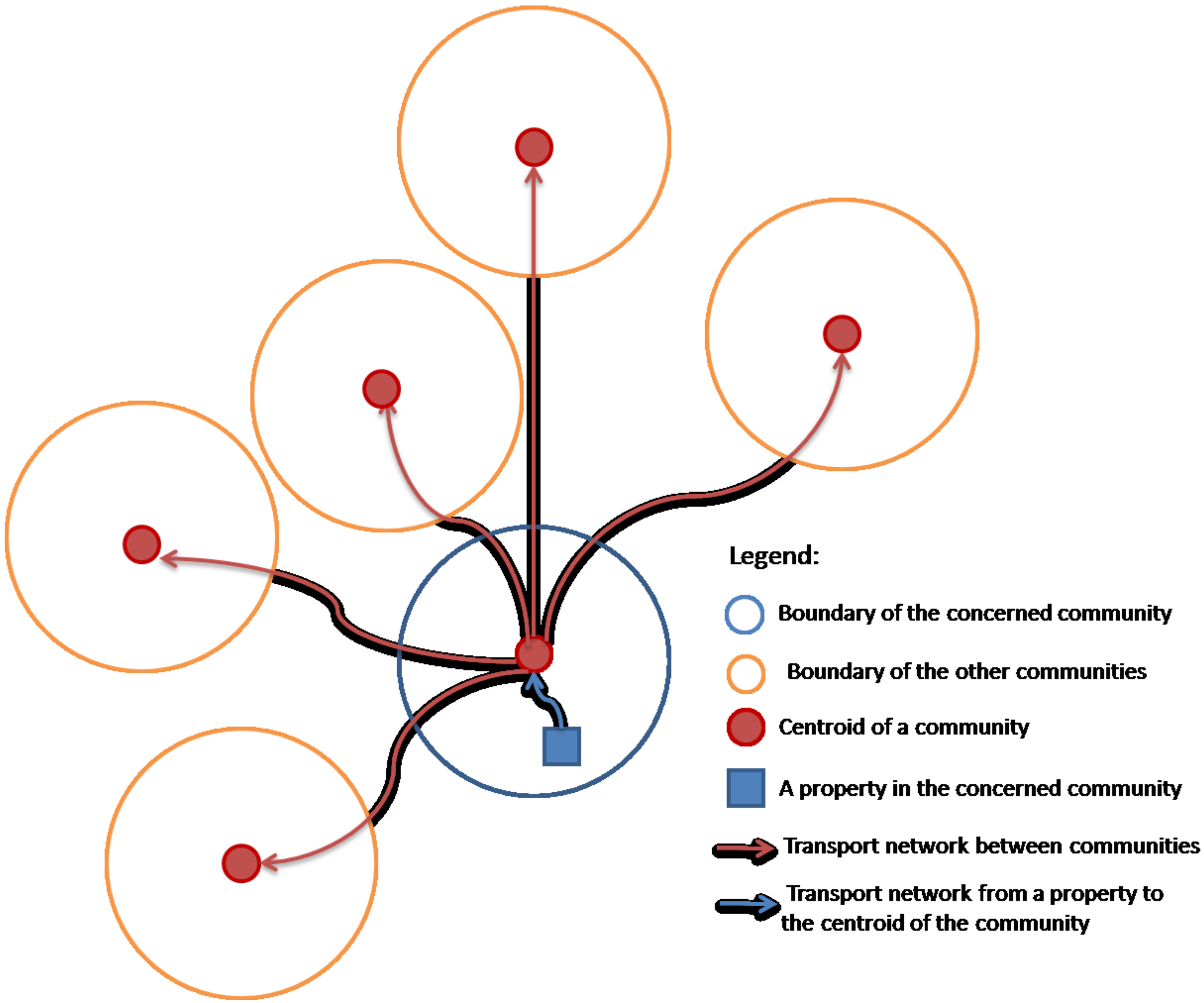
In summary, while the first and the second measures test for the metro effect on property values depending on the distance of a property to one or a number of metro stations, the third measure tests for the effect that the metro has on property values, based on the overall travel times and costs of using private and public transport modes (i.e. provides a network wide effect of the metro). The results using the three variables complement each other.

Figure 6-1: Illustration of the first and second accessibility measures offered by the metro (source: self-produced graph)



In the first measure (binary variable), the blue and the orange properties take a value of 1. For the second measure (accessibility to a number of metro stations), the blue properties are close only to one station and therefore take a value of 1, whereas the orange properties are close to two stations within the defined catchment area, hence take a value of 2. All properties outside the catchment area (i.e. the black properties) take a value of zero in the first two measures.

Figure 6-2: Illustration of the calculation of the value of the GC of travel (source: self-produced graph)



6.3 Land and property data

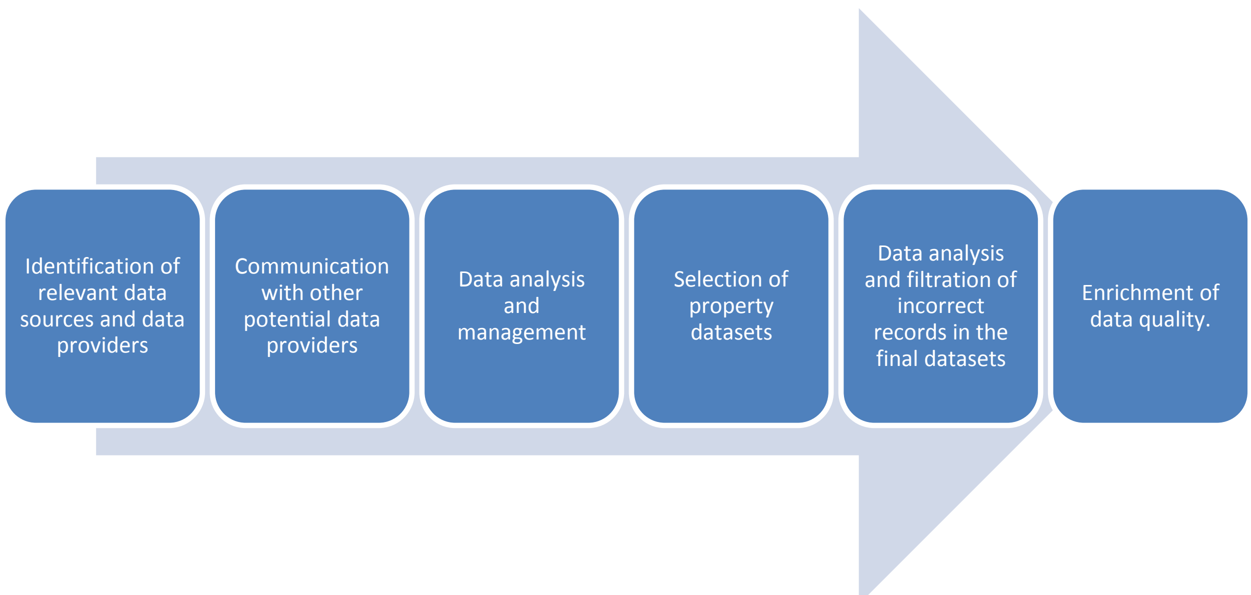
Communication was initiated with various authorities in Dubai to obtain suitable data on land and property records, for the period before the opening of the metro until the year 2011 (the end date for data collection in this study). Figure 6-3 **Error! Reference source not found.**illustrates the process of obtaining and completing the datasets to estimate the effect of the metro on land or property values. As a first step, communication was initiated with those authorities that were judged as likely to have records on the sale or rent of land and property in Dubai; these are RERA, a number of real estate agents such as Hamptons and Better Homes, large scale developers such as Nakheel and Emaar as well as DSC.

The initial contact revealed that that not all sale transaction records were automated and available in Dubai, as it was not mandatory by law to record these transactions until the year 2012, nevertheless RERA holds a number of these records from the year 2007 that could be used. In addition, a limited number of historic records were available in hard copy and located in archived folders. It was also found that a relatively small number of rent records were available through RERA. In addition, one real estate agent, named REIDN, gathers large datasets on almost all asking values from large scale developers and other real estate agents. Finally, a small number of rental values were available through the DSC.

After obtaining records on land and property data following the first contact with relevant authorities, a few entities declared that a number of hard-copy rental agreements were available through various sources in Dubai. For example, the Land Department holds some hard-copy rent agreements, Dubai Municipality holds cases of property rent disputes and Dubai Electricity and Water Authority (DEWA) holds copies of residential rental contracts, albeit only at the time at which the property is transferred to a new tenant. These entities were contacted in order to try to obtain automated data, but it was founded that the data was not automated and was not able to be provided for this study. In total 931,086 records were obtained from RERA, REIDIN and DSC.

Next, the obtained datasets were checked for accuracy and the records with missing attributes were removed. This was followed by a review of the suitable datasets in terms of their ability to provide unbiased estimates of the effect of the Dubai Metro on values. This review is detailed further later in this section. Although some limitations were found in the provided datasets in respect to a number of the attributes that probably affect land and property values, the effect of these limitations was ameliorated by manually enriching the datasets through either online sources or by calculating the variables using an Arc GIS tool.

Figure 6-3: A flow chart of the process of obtaining land and property data (source: self-produced graph)



This section begins by describing the obtained land and property datasets and discusses their advantages and limitations in light of the study design rules set out in chapter 5. This is followed by summary statistics of the selected datasets. Next, the available and the added missing variables that can affect property values in Dubai are listed.

6.3.1 Description of land and property datasets

This section discusses the obtained land and property datasets with the aim of selecting the datasets that are suitable for further analysis. There are a number of different sources of land and property data in Dubai. The Real Estate Regulatory Authority (RERA) provides records on land

and properties. RERA is the most trusted entity for providing land and property observations since it is an audited government body established for this purpose. Nonetheless, RERA was only established in the year 2007, and it is only since 2012 that land and property owners have been obliged to register with RERA and record all transactions. Before that, registering transactions was optional and only hard copies were kept in the Dubai Land Department. As a result, the number of observations that RERA holds until the year 2011 is relatively limited.

Real estate agents and commercial vendors also provide land and property data and examples of these are REIDIN, Hamptons, Better Homes and Harbor Real Estate. REIDIN, in particular, gathers a large database of repeated cross-sectional data for the sale and rental listings of residential and retail properties. They collect from other real estate agents, developers' websites and other property listings websites. The dataset from REIDIN, however, although larger in volume and more varied than any other, contains information only from 2009 onwards (around one year before the opening of the majority of the metro Red Line stations).

Other sources that hold records of transactions and listings of land and property in Dubai are the large local developers. The majority of their data, however, also forms part of the data provided by other sources like RERA and the real estate agents. Another source is DSC, which is a government entity that provides panel data, albeit for a limited sample size of rent transactions for over 5 years (2007 - 2011). For the purposes of this study, therefore, data was only gathered from RERA, REIDIN and DSC.

Table 6-1 is a summary of nine original property datasets in terms of land or property type, transaction years and the number of observations, as a whole, and in groups. Two sets of data are presented: original and filtered data. The original data presents the sample sizes for all obtained properties without clustering them to repeated sale groups, whereas the filtered data only considers the records of properties that were available at least once before and once after the opening of the metro.

It is worth mentioning that the sample sizes for the original datasets in the table are the reduced sample sizes after filtering for missing and incorrect observations. These include missing land or property values or the absence of other covariates; such as if the transaction value or the property area is recorded as a null figure. Incorrect records were also excluded; such as if the observation was recorded as a villa whereas the building name associated with the record refers to a tower of flats. Similarly, records were excluded which list a property under a building that is located in a different community than the one mentioned in the record.[19]

As discussed in the previous chapter, the aim is to test for the effect of the metro on the value of treated and control properties before and after the opening of the metro. The available records pre-2010, and both in and after 2010, are therefore presented. The number of available land and property records within the maximum selected catchment zone (1.5 km), compared to the records for further distances is also presented. It was found that, although the sample sizes in many datasets seem reasonable for analysis, the number of treated and control land or properties pre- and post-opening is not always sufficient to produce a consistent estimate for the effect of the metro on property values. In this regard, we have selected a threshold of 30 observations in each category and only datasets that meet this threshold for each category are included. The next subsections describe the obtained and the selected datasets from each of the data sources.

[19] The location of a building within Dubai (i.e. the community in which it is situated) was obtained from developers and real estate websites.

Table 6-1: A summary of the original and filtered RERA, REIDIN and DSC datasets

Data source	**RERA – Transactions**						**REIDIN – Asking values**		**DSC - Transactions**
Property type	Sale - Residential properties	Rent - Residential properties	Sale – retail properties	Rent - retail properties	Sale - Residential land	Sale – commercial land	Sale - Residential properties	Rent - Residential properties	Rent - Residential properties
Data type	Repeated cross-sectional data								Panel data
Panel years	2007-2011	2009-2012	2007-2011	2009-2012	2007-2011	2007-2011	2009-2011	2009-2011	2007-2011
Original data									
No. communities	**15**	**23**	**11**	**22**	**72**	**76**	**66**	**71**	**31**
No. observations	**50,070**	**8,918**	**7,495**	**4,581**	**2,301**	**2,334**	**441,807**	**415,079**	**835**
Treated (within 1.5km)									
No. observations up to 2009	8,882	47	1,062	241	290	313	68,095	77,102	459
No. observations after 2009	22,277	6,282	4,335	2,922	74	68	244,929	206,891	306
No. parcels	**145**	**11**	**93**	**146**	**267**	**591**	**566**	**565**	**153**
Control (beyond 1.5km)									
No. observations up to 2009	4,156	7	1,449	18	1,425	1,403	25,744	28,362	42
No. observations after 2009	14,755	2,582	649	1,400	512	550	103,039	102,724	28
No. parcels	**1,871**	**5**	**26**	**42**	**1,310**	**1,269**	**840**	**654**	**14**

Filtered data									
No. communities	**8**	**7**	**7**	**9**	**37**	**21**	**11**	**9**	**31**
No. observations	**39,308**	**1,523**	**3,419**	**2,311**	**320**	**343**	**165,978**	**81,248**	**835**
Treated (within 1.5km)									
No. observations up to 2009	4,562	44	711	241	35	5	43,436	22,362	459
No. observations after 2009	14,216	817	799	1,387	29	6	101,893	43,668	306
No. parcels	**101**	**9**	**42**	**9**	**37**	**28**	**37**	**28**	**153**
Control (beyond 1.5km)									
No. observations up to 2009	5,697	3	1,428	18	144	229	5,655	7,268	42
No. observations after 2009	14,749	659	481	665	112	103	14,933	7,950	28
No. parcels	**1,012**	**3**	**8**	**5**	**109**	**88**	**109**	**88**	**14**

6.3.1.1 RERA datasets

RERA provides repeated cross-sectional data for six types of actual land and property transactions for a period of over 5 years: sale and rental of residential properties, sale and rental of retail properties and sale of residential and commercial land. As the above table shows, the largest dataset from RERA is for records on dwellings (more than 50 thousand observations between 2007 and 2011). It is also evident that the number of observations in the period after the opening of the metro increases due to the greater effort invested by RERA in improving its documentation system as it became more established.

The data contains about three times as many treated parcels as control parcels, and around 74% of the total number of records is from after the opening of the Red Line.[20] The data is also spread over fifteen communities in Dubai because the most readily available and consistent data on property transactions are for the newly developed communities (i.e. developed after the year 2000) and most of these new communities are close to the metro Red Line. There are also limited sale transactions for properties located in all other communities in Dubai.

After data refinement, the number of properties in the sample reduces to around 39,000, distributed in eight communities. It should be noted that grouping observations to pseudo panel cohorts considerably reduces the number of available property observations before and after the metro operations. Properties in this dataset are scattered along the metro Red Line or are located far away from any metro station. Almost half of the records in the refined dataset are treated.

Considering now rental transactions relating to residential properties, the RERA dataset contains observations only from the year 2009 onwards. There are just fewer than 9,000 records of rented dwellings in this dataset and about two thirds of these are properties located within 1.5 km of a metro station. Refining this further to available cohorts in and after 2009, there are almost no

[20] A parcel is a piece of land the size of one building.

records of control observations and a rather small number of treated properties before the operations. This prevents the use of this dataset for analysis.

As for the records relating to the sale of retail properties, a reasonable sample size in treated and control groups was obtained after data filtration which allows this dataset to be considered further. It was found that there are more consistent sale records of retail properties in the recently developed mixed-use land communities in Dubai than there are in the older parts of the emirate. Nevertheless, the dataset containing rental records of retail properties is problematic due to the limited number of control properties before the opening of the metro and hence it is excluded.

RERA datasets on residential and commercial land contain records on a fair number of observations (above 2,300) with a wide spatial distribution (covering over 70 communities). After data refinement, however, it was found that 80% of the residential land transactions and 97% of the commercial transactions were for properties located more than 1.5 km from a metro station. The limited sample size for treated land and the large distribution of records on land located in control communities compared to that for treated communities prevents the further use of these two datasets.

6.3.1.2 REIDIN datasets

REIDIN provides two datasets (each containing over 400,000 records) on sale and rent asking values of residential properties from the year 2009 to 2011. The property observations in the sale and rental datasets are spread across 66 and 71 communities, respectively. It should be noted that there is a much larger number of records available for the years 2010 and 2011 compared to those for 2009. This is mainly due to the fact that the data collection effort increased with time.

After filtration for clusters of properties that do not repeat over time, a considerably reduced, but still very large, number of observations remained (more than 165,000 sales and 81,000 rents).

There also remained substantial sample sizes for treated and control properties (above 5,600 in each case). The sale and rental listings are scattered along different parts of the metro Red Line, however, or are located far from any metro station. The corrected datasets contain a much smaller number of communities. The sale records cover essentially the same communities as the rental records, although the former set contains two additional neighbourhoods.

Although the REIDIN datasets contain records from the year of the metro's inauguration (2009), it is still possible to use this dataset for two reasons. First, the year 2009 is selected as a pre-treatment year (see the explanation in the previous chapter). Second, the filtered property records do not include properties within the vicinity of the ten metro stations opened in that year. Although it is recognised that this adds a limitation to the results generated using this dataset, since it only considers one year before operations started, REIDIN nonetheless provides a different type of property record compared to the RERA datasets (i.e. listings versus transaction records), which is worth considering.

6.3.1.3 DSC datasets

In contrast to the repeated cross-sectional data provided by the other two data sources, the Dubai Statistics Center provides five-year panel data on rented residential properties, albeit for a small number of properties (835 observations). The records are from the year 2007 to the year 2011 and are spread over 31 communities. As this dataset is a panel data it is not necessary to refine it any further.

Although the main advantage of this dataset is that it is genuine panel data, containing records for almost all residential communities in Dubai, it covers a limited number of properties per community (on average five properties). As explained by the DSC, only one property per building is considered and these are selected without any pre-study of the choices made. As a result, the choice of these properties or buildings can be questioned. In addition, 91% of properties in this dataset are close to a metro station. This leaves quite a small number of control properties. For these reasons, this dataset is not considered further in the research.

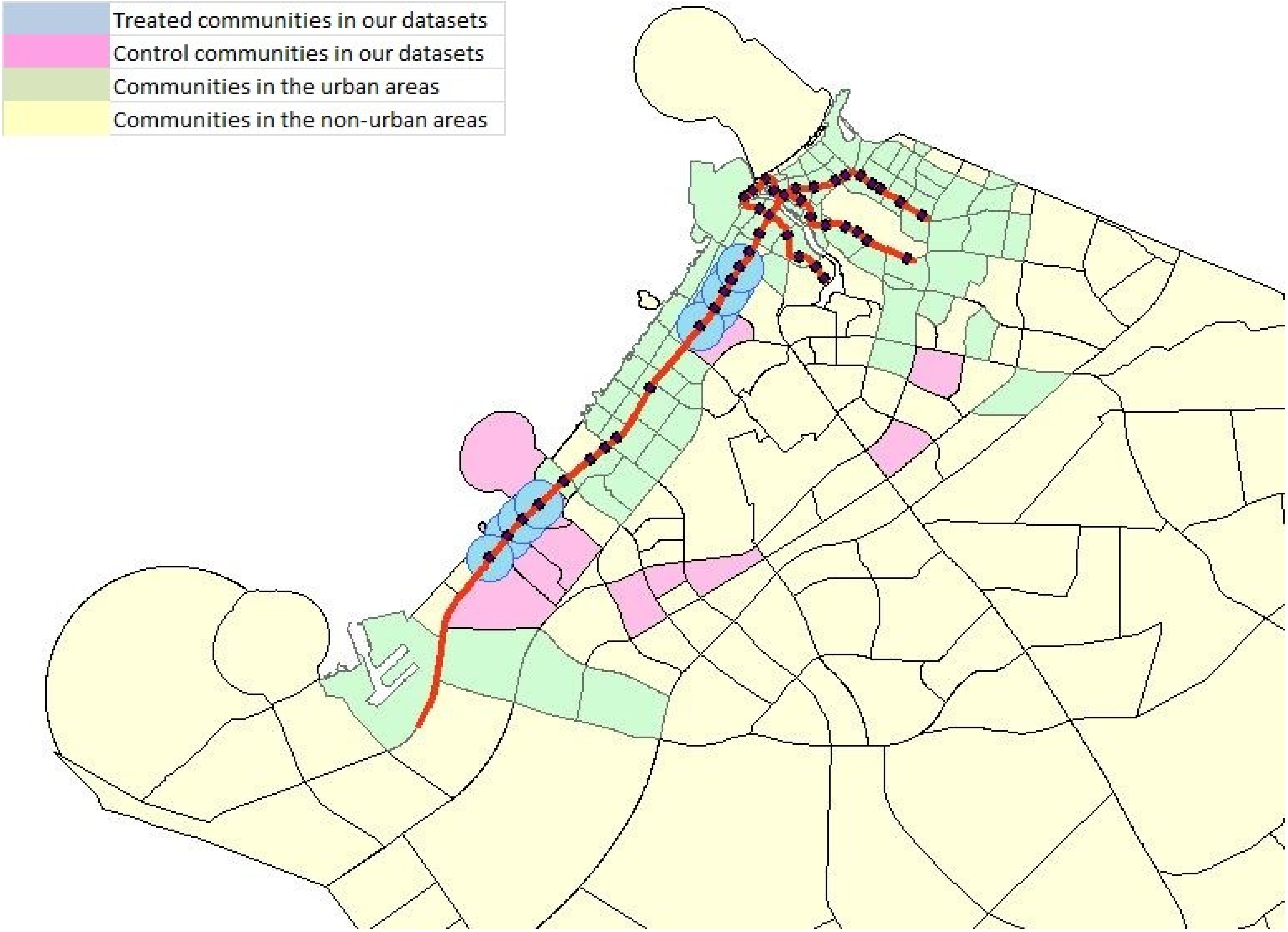
6.3.1.4 Overall summary

From the nine original datasets provided by RERA, REIDIN and DSC, four are suitable for further analysis; these are records on sale transactions of dwellings and retail properties (RERA) and the sale and rental listings of dwellings (REIDIN). The difference in the type of records (transactions versus listings) makes it possible to compare the effect of the metro on the actual value of properties using RERA datasets and the perceived value of properties using REIDIN datasets.

Figure 6-4 presents the geographical distribution of the obtained property records in the selected datasets as well as the urban area in Dubai. It should be noted that the sample data covers only a part of the urban area, albeit for different types of property records. It is also acknowledged that the spatial coverage of the datasets is limited and this has the following implications. First, the results of this study are estimates for the effect of the Dubai Metro on a sample of properties in Dubai that are located in the newly developed communities (the reader is referred to chapter 4). Second, since observations are available either in close proximity to the Red Line or in control communities, this study tests for the effect of the metro Red Line only. Nevertheless, this study is the first attempt to examine the effect of the recently opened Dubai Metro on property values and, given that all datasets contain records for properties located in the treated and control communities pre- and post-operations, it remains possible to obtain consistent estimates and reduce the bias from estimation models.[21]

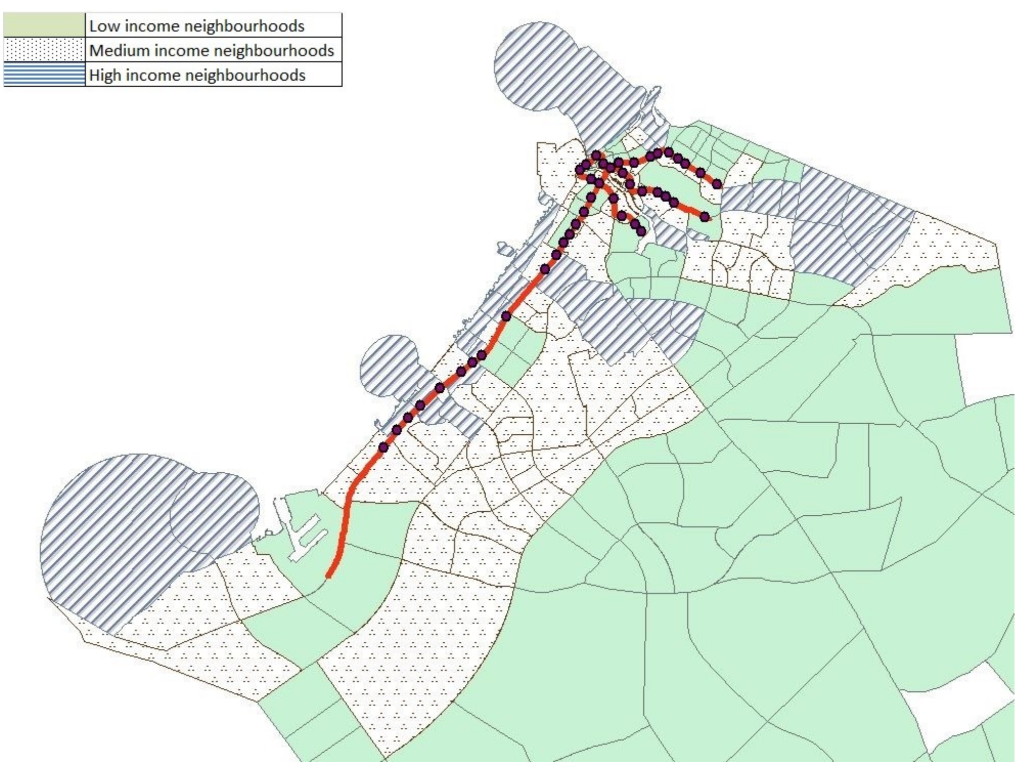
[21] The number of properties in each dataset compared to the number of properties in a community is presented in appendix E.

Figure 6-4: Communities that contain property data in the sample used and communities in the urban and non-urban area of Dubai as of the year 2011



One issue to consider is whether the choice of the metro route and the location of stations may have been biased towards specific neighbourhoods. For example, one may question whether the metro serves areas with specific household income levels or property values. If this were the case, it could lead to reverse causality between accessibility to the metro and property values. With the limited available information (income values at a property or community level are not measured, so only an indication on the income level is available[13] (low, medium, high)), Figure 6-5 maps the income level of residents. It can be observed from this that the metro serves communities with all types of income levels, therefore it can be reasonably assumed that the income level (and, as an extension, property value) is exogenous to the choice of metro location. As a consequence, no feedback effect is expected between property value and the Dubai Metro.

Figure 6-5: The Red and Green metro lines and an indication of community income level (source: self-produced graph based on the classification of income levels)



There may, however, be other fundamental differences in the characteristics of the neighbourhoods served by the metro (treated) compared to neighbourhoods not served by the

[13] The classification on income levels is based on the rental indices for residential properties in a community and the indication was selected in association with Dubai Real Estate Regulatory Authority.

metro (control). Table 6-2 **Error! Reference source not found.**therefore provides summary statistics for the available variables (population density, employment density and shopping commuters). Since the most disaggregated densities are at a community level based on the sample of property observations obtained for this study,[14] figures are provided separately for all treated and control communities in Dubai and for the treated and control communities that contain property observations in our datasets.

Table 6-2: A summary of the characteristics of the treated and control communities

Variable	Treated communities		Control communities	
	Mean	**Standard deviation**	**Mean**	**Standard deviation**
All communities in the urban area of Dubai				
Population density*	16,600	25,500	3,700	6,200
Employment density*	8,100	13,200	1,300	3,200
Shopping commuters**	1,543	4,344	937	1,975
Communities in the datasets used in this study				
Population density*	2,300	2,600	1,600	1,200
Employment density*	1,300	1,200	100	200
Shopping commuters**	1,138	1,598	578	1,079

* These are measured as population or employment number per square kilometre.

** This is measured as the number of shopping commuters to a community in a peak hour.

Table 6-2 suggests a significant difference in the population densities of individual treated and control communities among all neighbourhoods in Dubai, but there is no substantial difference between the means of the population densities in the treated and control communities in our sample. On the other hand, a substantially higher employment density, and a greater number of shopping commuters is evident in treated communities compared to control communities, both in

[14] For more information on the obtained property datasets, the reader is referred to chapters 5 and 6.

the sample used here and for all Dubai. As a result, these variables need to be controlled for in the proposed models.

It is assumed that there is no impact of the metro on the distribution of land use in Dubai during the study period (2007-2011) (i.e. there is no feedback of the metro on the population, employment and shopping densities). This assumption is based on the land use distribution and developments having remained unchanged during that time (Dubai Municipality, 2010). Due to the increased supply of offices that had already been planned, however, employment density has increased with time, and this needs to be considered.

6.3.2 Summary statistics of the final datasets

In this section, we provide summary statistics for the final selected property datasets. Table 6-3 presents the sample sizes for the three main catchment areas, as discussed in the previous chapter (0.5 km, 1 km and 1.5 km), and for the two data structures. The sample size for the pseudo panel data is based on the 'criteria option 4' as discussed in section 5.3. It should be noted that the listing datasets contain a larger number of observations compared to that in the transaction datasets. This is expected, since there are generally more properties listed for sale or rent than there are actual transactions. The property data sample sizes used in this study are in a similar range to those used in previous such studies, and are considered large enough for analysis.

Comparing the average property value per unit area in the repeated cross-sectional data and the pseudo panel data (Figure 6-6 to Figure 6-9), it can be seen that the average values are more similar when the metro catchment area is defined at 1.5 km from a station, followed by the 1 km and 0.5 km distance bands. As the catchment area increases, the number of records within that area also increases, which results in more comparable values using the two data structures.

Table 6-3: Summary statistics of the datasets at different catchment areas (0.5 km, 1 km and 1.5 km)

Data source	RERA	REIDIN	REIDIN	RERA
Data type	Sale transactions of residential properties	Sale listings of residential properties	Rental listings of residential properties	Sale transactions of retail properties
Panel years	2007-2009 (before metro), 2010–2011 (after metro)	2009 (before metro), 2010-2011 (after metro)	2009 (before metro), 2010 and 2011 (after metro)	2007-2009 (before metro), 2010-2011 (after metro)
Coefficient of variation in property value	1.5 (rcs) 0.76 (pp)	0.98 (rcs) 0.57 (pp)	0.57 (rcs) 0.62 (pp)	1.12 (rcs) 1.49 (pp)
Number of observations in the repeated cross-sectional data*				
Total number	**39,308**	**165,978**	**81,248**	**3,419**
Properties within 0.5 km of a metro station	7,911 (20%)	20,914 (13%)	7,261 (9%)	708 (21%)
Properties beyond 0.5 km of a metro station	31,397 (80%)	145,064 (87%)	73,987 (91%)	2,711 (79%)
Properties within 1 km of a metro station	21,326 (54%)	107,520 (65%)	49,254 (60%)	1,237 (36%)
Properties beyond 1 km of a metro station	17,982 (46%)	58,458 (35%)	31,994 (40%)	2,182 (64%)
Properties within 1.5 km of a metro station	28,959 (74%)	116,828 (70%)	51,618 (64%)	1,510 (44%)
Properties beyond 1.5 km of a metro station	10,349 (26%)	49,150 (30%)	29,630 (36%)	1,909 (56%)
Number of observations in the pseudo panel data*				
Total number	**3,344**	**2,288**	**823**	**336**
Cohorts within 0.5 km of a metro station	223 (7%)	428 (19%)	154 (19%)	75 (22%)
Cohorts beyond 0.5 km of a metro station	3,121 (93%)	1,860 (81%)	669 (81%)	261 (78%)
Cohorts within 1 km of a metro station	573 (17%)	1,407 (61%)	435 (53%)	171 (51%)
Cohorts beyond 1 km of a metro station	2,771 (83%)	881 (39%)	388 (47%)	165 (49%)
Cohorts within 1.5 km of a metro station	891 (27%)	1,674 (73%)	493 (60%)	280 (83%)
Cohorts beyond 1.5 km of a metro station	2,453 (73%)	614 (27%)	330 (40%)	56 (17%)

Legend: RCS: repeated cross-sectional data. PP: pseudo panel data

* The cells contain the number of properties, while the figure in brackets is the percentage of observations to the total sample size.

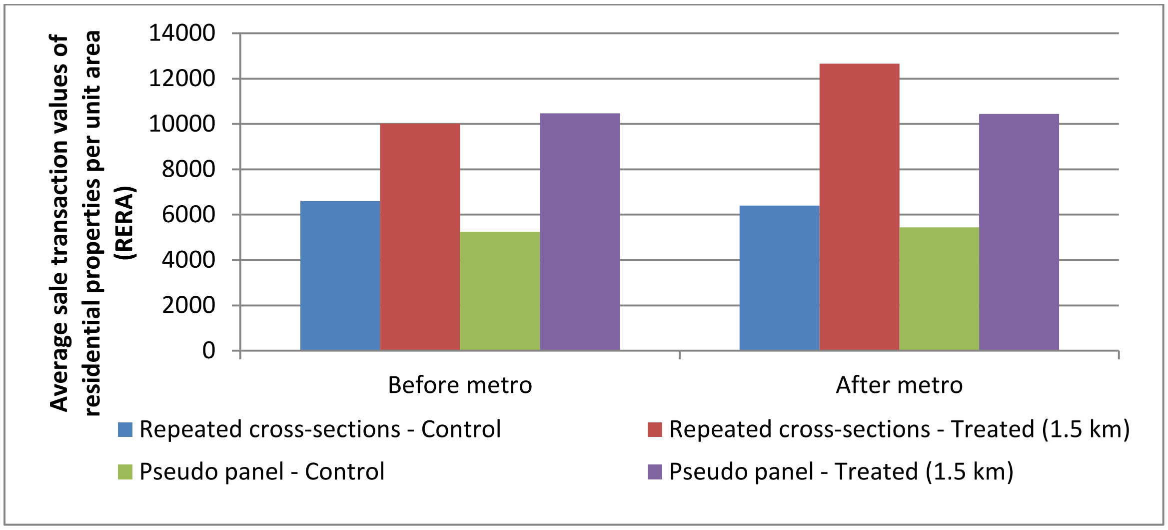
In addition, the difference in the average value per unit area within the same property category (treated versus control) using the repeated cross-sectional data versus the pseudo panel data is larger in the residential datasets compared to the retail dataset. In line with these figures, the coefficient of variation for property values (Table 6-3) is also more comparable between the two data structures for the retail dataset compared to the residential datasets.

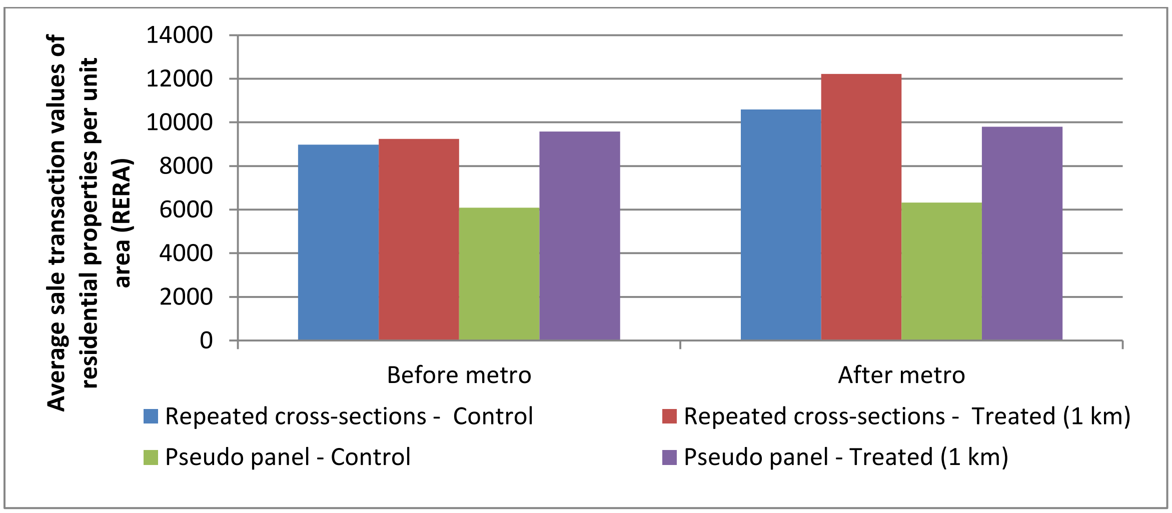
The structure of observations by catchment area (0.5 km, 1 km, 1.5 km) between repeated cross-sectional data and pseudo panel data compares better between type of data for REIDIN than for RERA datasets. An explanation for this is that REIDIN provides a large sample size, and therefore a larger number of observations are grouped in a cohort compared to the RERA residential dataset. This results in more similar average property values between the two data structures.

The cohort sizes in the pseudo panel data are also plotted for each dataset (presented in appendix D). While for the RERA residential dataset, the majority of cohorts contain less than 50 observations, it should be noted that 80% and 50% of the cohorts in the REIDIN sale and rental datasets, respectively, contain around one hundred records. For the retail dataset, the cohort sizes are relatively smaller; this can be related to the smaller number of actual retail properties in a cohort.

The process of constructing pseudo panel data results in a great deal of information being lost, especially when the cohort size is limited and the heterogeneity across properties within a group is large. A greater difference in the average property values between the repeated cross-sectional data and the pseudo panel data indicates a larger measurement error in the latter data. This in turn can lead to biased estimates. The results presented in chapters 7 to 9 provide and compare the estimates using both data structures, with the differences due to the measurement error being explained in more detail in chapter 10.

Figure 6-6: Average sale transaction values of residential properties per unit area, for different catchment areas, before and after the metro using the repeated cross-sectional data and the pseudo panel data





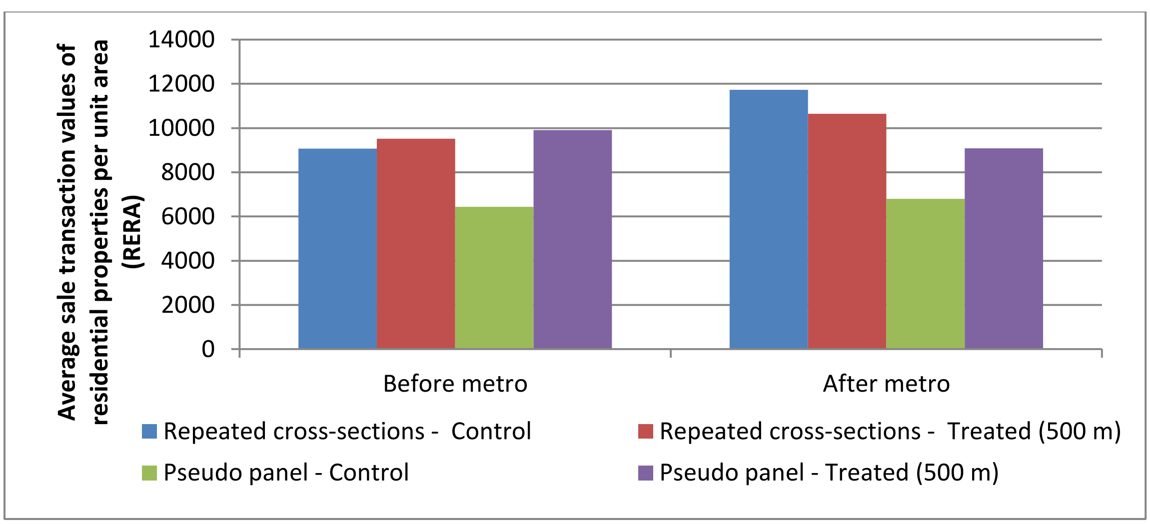


Figure 6-7: Average sale listing values of residential properties per unit area, for different catchment areas, before and after the metro using the repeated cross-sectional data and the pseudo panel data

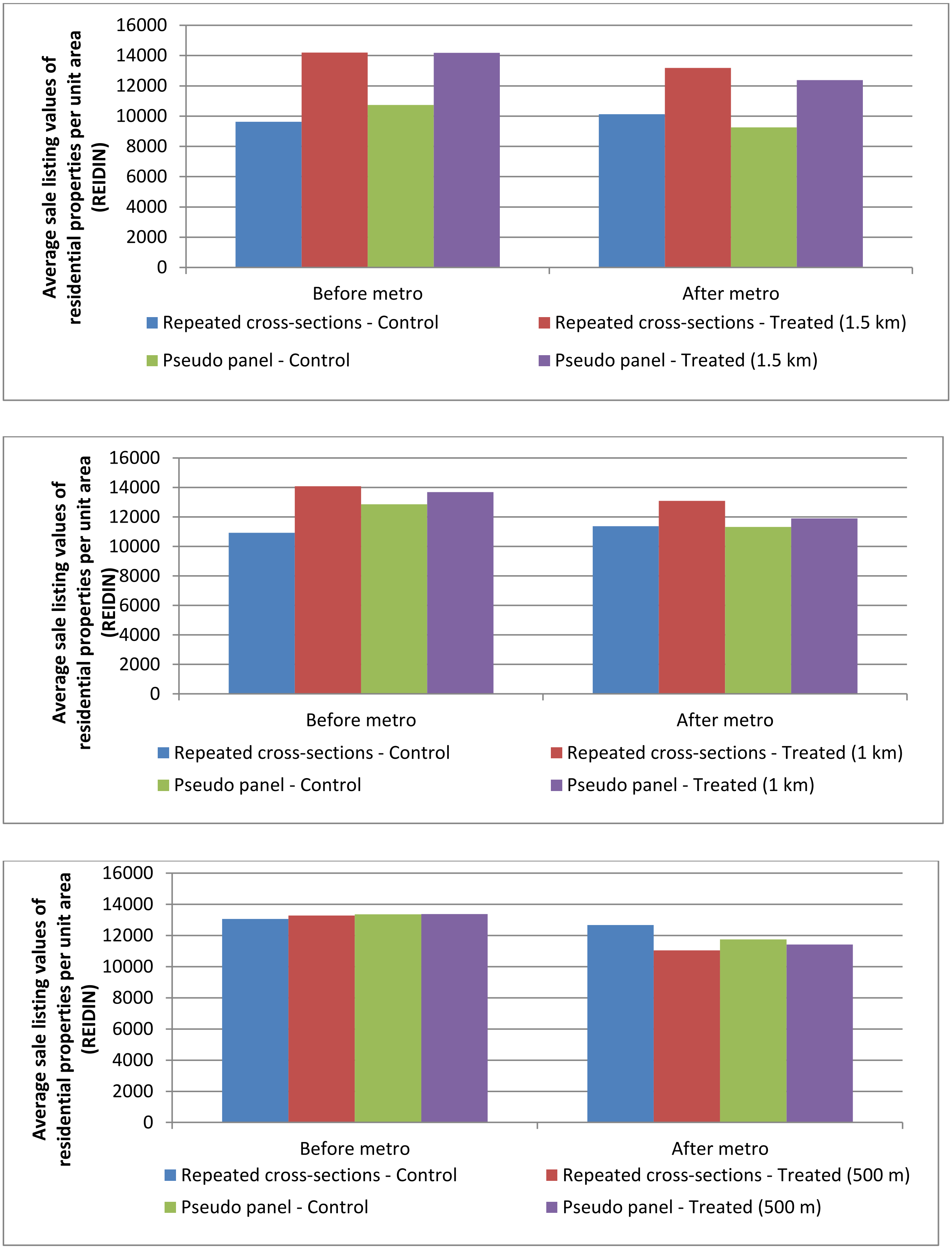
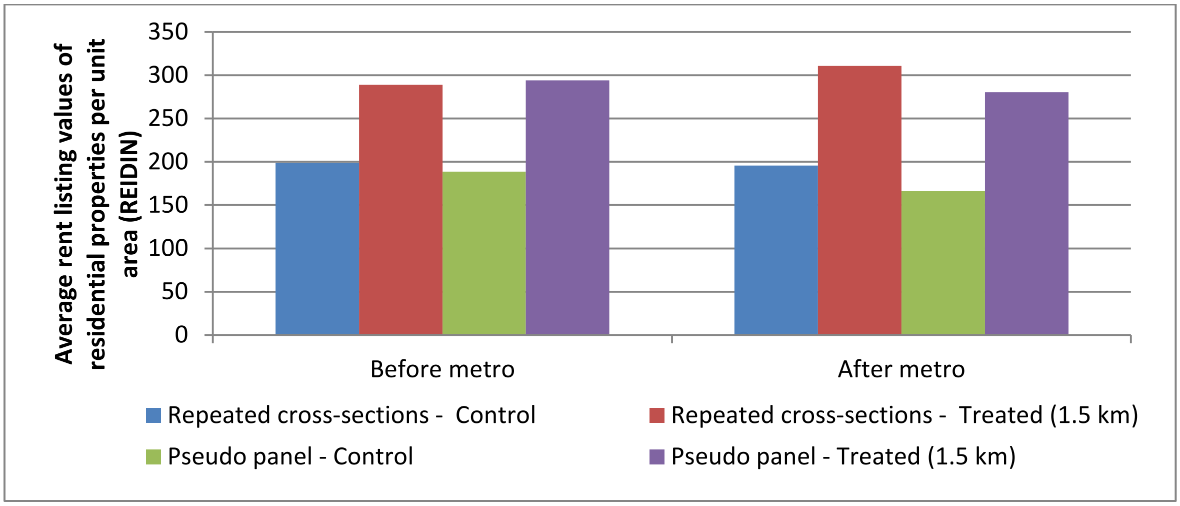
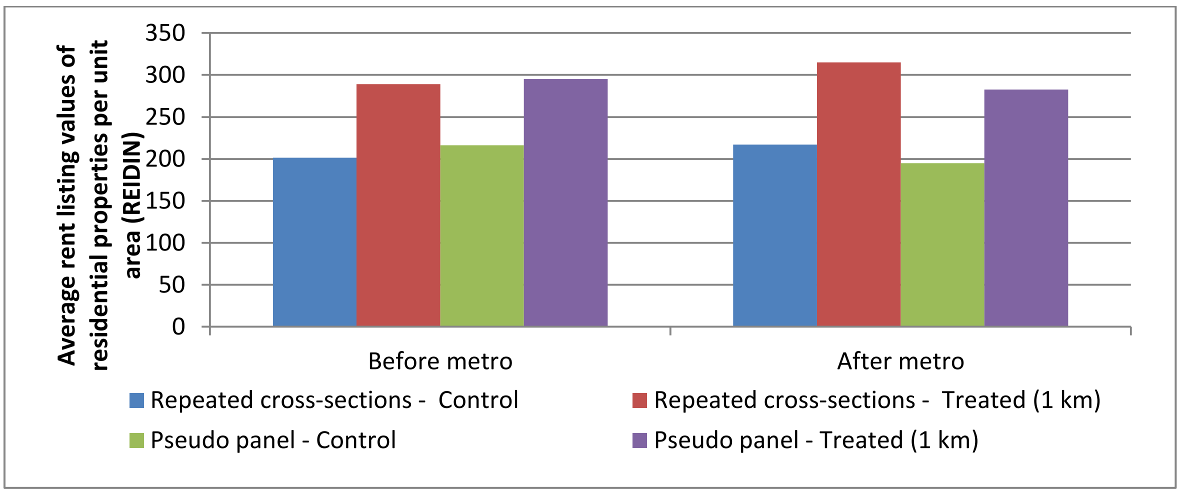


Figure 6-8: Average rental listing values of residential properties per unit area, for different catchment areas, before and after the metro using the repeated cross-sectional data and the pseudo panel data





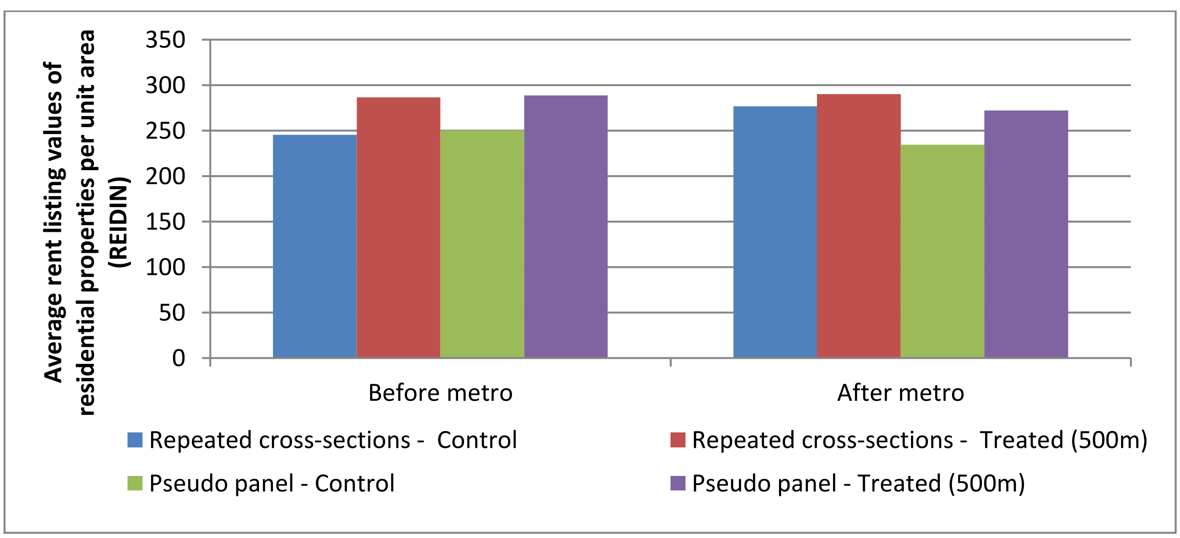
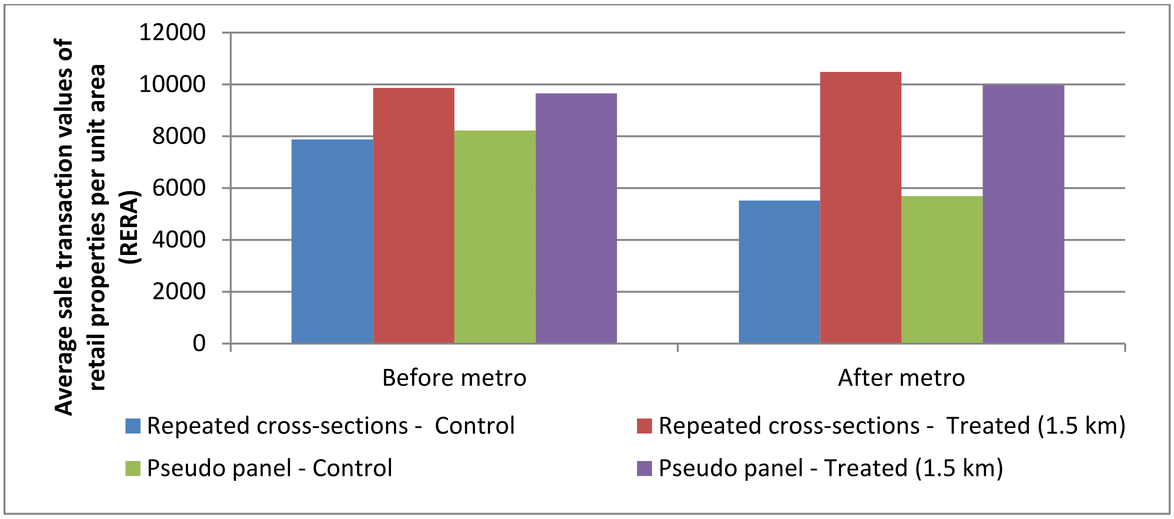
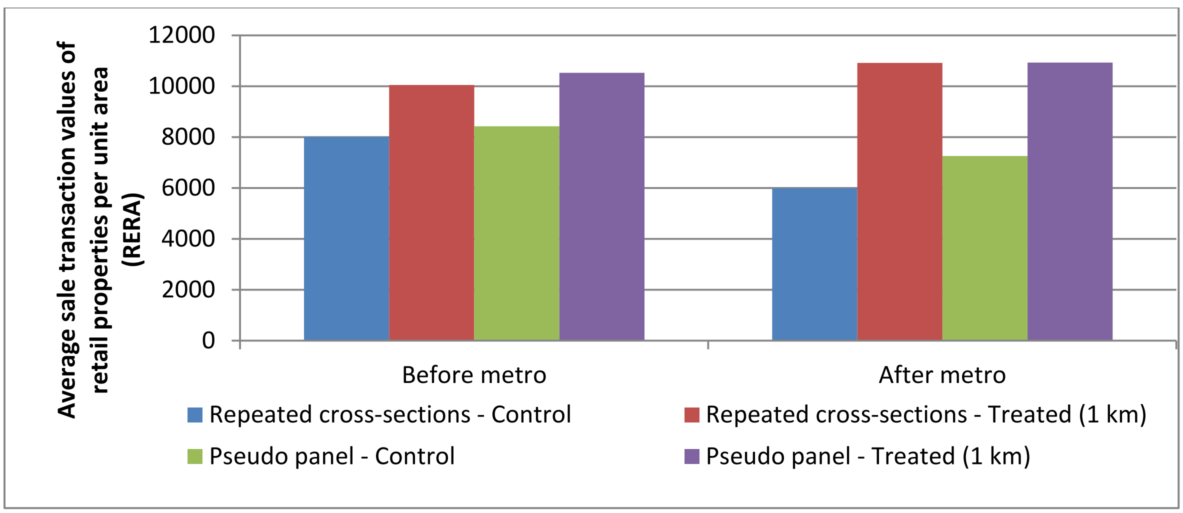
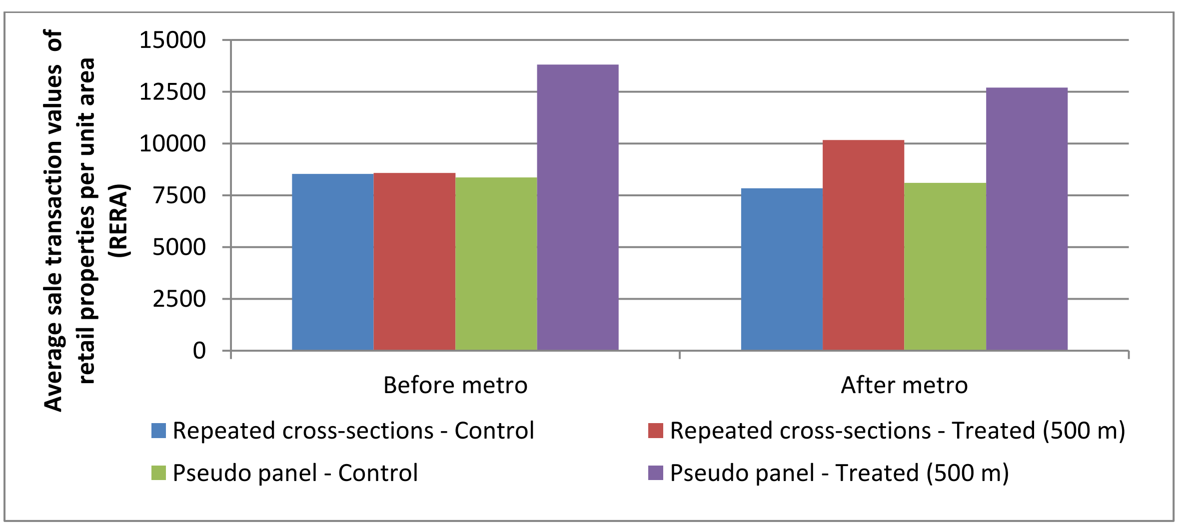


Figure 6-9: Average sale transaction values of retail properties per unit area, for different catchment areas, before and after the metro using the repeated cross-sectional data and the pseudo panel data







6.3.3 Provided and added attributes

Based on the literature review and the results of the meta-analysis, it was intended, ideally, to obtain data on as many covariates that might be expected to impact on property values as possible. This section reviews the obtained variables and explains how missing covariates were added. The selected datasets include the sale or rental value of a property, the year of transaction or listing, an indication of the property location, the community in which a property is situated, property area and, for the REIDIN datasets, also the number of bedrooms, the age and other building attributes. Table 6-4 lists the covariates related to this study and presents the observed and added variables in each of the selected datasets. The covariates are divided into the three main factors as presented in chapter 2 (i.e. internal factors to the property, external factors and economic factors).

Starting with internal characteristics, property type is distinguished (i.e. residential versus retail) and the attributes related to property size are obtained. In addition, it was attempted to obtain as many building characteristics as available, either from the data sources or by adding them manually using web-based sources on property listing websites (for which it was assumed that the information provided online was sufficiently accurate). The original RERA datasets contain a limited number of physical characteristics of a property. To account for the missing property characteristics, as explained in the previous chapter, the repeated cross-sectional data allows similar records to be clustered together in order to control for the unobserved heterogeneity at a cluster level. The REIDIN datasets, on the other hand, contain the number of bedrooms. In addition, the building name in the REIDIN datasets makes it possible to obtain the building characteristics (age, the availability of a gym, a swimming pool and a porter service).

Moving to the external factors, and specifically the transport variables, the distance to the nearest metro station was obtained, the number of accessible metro stations within the specified catchment areas, the distance to the nearest highway available between 2009 and 2011, the generalized cost of travel by public and private transport, with and without the metro, as well as the modal split between private and public transport before and after the opening of the metro.

While the RERA datasets report the parcel ID as the indicator of the exact property location within Dubai, REIDIN reports on the building name that contains the property. The location indicator then makes it possible to find the main attractions in close proximity to a property. This is achieved either through using an ArcGIS software that contains land use and transport network data to measure the distance between a property and transport links or other amenities (if the parcel ID is available), or through using online sources that report the distance between a building and the major attractions in the emirate (if the building name is available). Access to the surrounding amenities can also affect property values and, therefore, covariates have been manually included to measure the proximity to major attractions (schools, hospitals and shops, etc.). All distances are straight line distances, for reasons explained in section 5.1. The community level employment density and the number of shopping commuters to a community were also included, since this is the most disaggregated data available in Dubai, and is required in the empirical models (the reader is referred to chapter 2). Finally, to measure the effect of time on the reported values, the observation year is considered.

Although we considered other variables that can affect property values, they were not added to the models for a variety of reasons. For example, consideration was initially given to adding the public transport share in the community where the property is located. This variable measures the preference of the residents towards the metro and can affect the bid value of a property. Nevertheless, this attribute is correlated with the measure of accessibility offered by the metro, and can lead to potential bias due to model over-specification, hence was eliminated. In addition, some studies have suggested that the specification of the metro service (e.g. its frequency) can affect property values (Ahlfeldt, 2013; Gibbons & Machin, 2005), however as the specifications were the same in all metro stations in Dubai and remained unchanged for the period of the analysis, these variables did not add value to the regression models.

Descriptive statistics for the contextual factors in this study are provided in Table 6-5 and distinguished per group (treated and control). As the aim is to describe the differences in contextual factors between residential and retail properties, and given that the distribution of property records across residential datasets are similar, average statistics are provided for all residential datasets (i.e. as the average for the sale transactions, sale listings and rental listings) as well as the statistics for retail properties. It can be seen, not surprisingly, that the mean area of residential properties is higher than that for retail properties. Also, while the distances from residential and retail properties to the nearest schools and hospitals are similar, the distance from retail properties to the nearest shopping area is smaller than that for residential properties. It can also be observed that the mean distance of retail properties to the nearest metro station is smaller than that for residential properties, although the number of retail properties in close proximity to more than one station within the predefined catchment area of 1.5 km is slightly less than that for dwellings. Finally, the table reveals that retail properties are closer to highways compared to residential properties. These variables are added to the estimation models, as appropriate, to control for the differences between treated and control areas and to estimate their effect on the value of properties.

Table 6-4: Covariates considered for the study

Factor	Type	Covariate description	Measurement unit	Source of information	RERA datasets	REIDIN datasets
Internal factors to property	Characteristics of the property	Area	Square metre (sq.m)	Property data sources		
		Number of bedrooms	Number	Property data sources		
	Characteristics of the building	Age	Years	Online sources		
		Availability of a gym	Dummy 1, 0 otherwise	Online sources		
		Availability of a swimming pool	Dummy 1, 0 otherwise	Online sources		
		Availability of a porter service	Dummy 1, 0 otherwise	Online sources		
External factors	Transport services	Distance to the nearest metro station	Kilometre (km)	ArcGIS or online sources		
		Number of metro stations within the catchment zones	Number	ArcGIS or online sources		
		Distance to the nearest highway	Kilometre (km)	ArcGIS or online sources		
		The generalized cost (GC) of travel using public transport	Currency unit (AED)	Raw data from RTA and then own calculation		
		The generalized cost (GC) of travel using private transport	Currency unit (AED)	Raw data from RTA and then own calculation		
		Share of public transport trips to total motorized trips	%	RTA		
		Share of private transport trips to total motorized trips	%	RTA		
	Location and community	Community name	Name	Property data sources		
		Property location ID	Parcel ID or building name	Property data sources	Parcel ID	Building name
	Access to the surrounding amenities	Distance to the nearest school	Kilometre (km)	ArcGIS or online sources		
		Distance to the nearest hospital	Kilometre (km)	ArcGIS or online sources		
		Distance to the nearest shops	Kilometre (km)	ArcGIS or online sources		
	Densities	Employment density	Number / sq.km	RTA		
		Shopping commuters	Persons per hour	RTA		
Economic factor	Year	Year dummies	Dummy 1, 0 otherwise	Property data sources		

Cell colour indication

	Data observed from the property data source
	Data collected from other sources and added manually by the authors
	Data not observed and cannot be measured

Table 6-5: Descriptive statistics of contextual variables in the selected datasets

	Treated residential properties				Control residential properties			
Variable	**Mean**	**Std. Dev.**	**Min**	**Max**	**Mean**	**Std. Dev.**	**Min**	**Max**
Property area (sq.m)	130	63	16	1579	213	174	16	3816
Number of bedrooms	1.75	0.88	0	5	2.44	1.59	0	7
Age (years)	2.96	2.11	-1	8	2.49	1.82	-2	8
Gym available	0.92	0.27	0	1	0.59	0.49	0	1
A swimming pool available	0.51	0.50	0	1	0.50	0.50	0	1
Porter service	0.97	0.18	0	1	0.66	0.47	0	1
Distance to the nearest school (km)	1.99	0.98	0.30	3.05	1.96	1.21	0.30	3.99
Distance to the nearest hospital (km)	3.74	1.31	1.51	5.72	3.96	1.25	1.51	5.72
Distance to the nearest shops (km)	5.62	1.90	0.60	8.89	5.59	2.34	2.51	8.89
Log employment density	6.7	0.7	3.9	7.7	5.4	1.6	3.7	7.6
Log shopping density	5.8	1.6	4.6	10.4	5.7	1.9	1.7	10.4
Distance to the nearest metro station (km)	1.14	1.65	0.11	1.50	4.54	3.66	2.61	11.9
Number of stations within 1.5 km	1.99	0.72	1	4	0	0	0	0
Weighted average GC using private and public transport (AED)	21.0	4.4	14.8	29.3	19.1	4.5	0.4	24.6
Distance to the nearest highway (km)	3.68	1.93	0.84	6.36	4.03	2.73	0.86	13.94

	Treated retail properties				Control retail properties			
Property area (sq.m)	140	67	14	1114	59	22	25	149
Distance to the nearest school (km)	1.99	1,13	0.1	3.99				
Distance to the nearest hospital (km)	2.97	1.56	0.79	5.60	4.31	0.31	3.22	4.40
Distance to the nearest shops (km)	0.95	0.45	0.08	1.97	0.37	0.58	0.20	2.49
Log employment density	5.9	1.6	4.0	8.3	3.8	0.1	3.7	3.9
Log shopping density	4.9	1.7	2.0	8.2	8.9	0.9	5.5	9.2
Distance to the nearest metro station (km)	0.82	0.35	0.19	1.37	5.01	3.44	2.35	10.00
Number of stations within 1.5 km	1.30	1.28	1	3	0	0	0	0
Weighted average GC using private and public transport (AED)	19.0	2.1	14.5	24.7	17.1	0.5	15.0	17.3
Distance to the nearest highway (km)	4.96	1.57	1.37	7.54	1.27	1.08	1.00	6.36

6.4 Conclusions and recommendations on data management

This chapter has presented the transport and land and property data as well as the attributes considered for the study. While the transport data was more easily obtained, the land and property data required a great deal of filtration and consistency checks. Datasets were chosen that contain observations by property clusters or pseudo panel groups in treated and control communities before and after the opening of the metro. As a result, four of the nine original datasets were found to be suitable for this research.

It was found that while RERA provides a panel of property data covering a longer duration (3 years before and 2 years after the treatment), REIDIN provides data for only one year and two years before and after the treatment, respectively, but for a much larger sample size. The relatively limited number of observations on land and property transactions and listings in Dubai was a result of the lack of enforcement of data collection before 2012. Overall, the property data provided from RERA is more reliable because actual transactions are provided, and also because this data covers a wider time span and the entity is audited annually. Nonetheless, the considerably larger datasets (in terms of sample size) provided by REIDIN are valuable and unique sources for records of sale and rental listings of residential properties.

The descriptive statistics for each data structure are also provided in terms of the data arrangement, the sample sizes, and the distribution of property values. By collecting and consolidating the obtained information from each data source and adding a new set of variables per property record, it has been possible to create unique and much richer residential and retail property datasets for Dubai. This can be used in future for other analyses.

There were some challenges during data collection, which mainly arose due to the scattering of data among different sources and the lack of proper documentation of land and property records in Dubai. First, historical records of transactions and listings of land and property are distributed among various authorities, each with its own recording method. Contacting these authorities and obtaining approvals for data collection, analysing each dataset, reporting back to them and consolidating a large number of land and property observations from the three main sources was challenging and consumed the bulk of the time spent at this stage.

Second, there are a number of missing attributes that affect land and property values in the original datasets which had to be enriched manually. This was done by manually gathering the missing information on some of the physical attributes of the building in which a property is located in, as well as collecting data on the connectivity of land and property to transport services and major amenities. Adding the extra information for a large number of observations consumed unexpected extra effort at the data collection stage. Nonetheless, the enhanced datasets contain a richer set of variables that are related to this study.

There were a few lessons learnt from the data collection stage that inform feedback to RERA as the entity responsible for collecting and managing land and property data in Dubai. Since historical land and property data is quite important for various studies and is not readily available, it is proposed that the available hard copy data is collected, automated and stored in an easily retrievable electronic form. RERA could potentially collect, consolidate and automate these valuable historical transactions on land and property data from relevant entities, such as DM, DWEA and Land Department. Although this project would consume a great deal of resources and time, it would provide valuable information for future studies.

In addition, it is proposed that a new law be enacted requiring all residents of Dubai to submit their details yearly (anonymously), including basic information on household structure (e.g. number of family members, income level, and age), exact residency location (the parcel ID), property attributes and the value of their property (as per the rental agreement or sale contract if applicable). This database would not only provide a valuable source of information for future studies but also helps to monitor the population distribution in Dubai.

Chapter 7. ESTIMATING THE EFFECT OF PROXIMITY TO A METRO STATION ON PROPERTY VALUES VIA A DID ESTIMATOR

7.1 Introduction

The difference-in-differences (DID) estimator is used to estimate the effect of a treatment or an innovation and has been widely adopted by researchers following its use by Ashenfelter and Card (1985) in a study identifying the impact of training schemes on earnings. DID compares the value of a group exposed to a treatment (treated) to a group unaffected by the treatment (control), before and after the treatment occurs. It produces an unbiased estimate for the effect of the treatment provided that the DID assumptions are met.

Although the use of the DID method is increasing, the number of previous studies that have used the DID to estimate the effect of a transport system on land and property values is still relatively small (e.g. Agostini and Palmucci, 2008; Ahlfeldt, 2013; Ahlfeldt and Wendland, 2009; Concas, 2012; Dubé et al., 2014; Gibbons and Machin, 2005; Wu, 2012). This chapter presents and discusses the estimates for the effect of proximity to a metro station on the value of residential and retail properties by means of a DID estimator.

The chapter is divided as follows. Section 7.2 introduces the basics of the DID method and section 7.3 develops the DID models for the study. The results are presented and discussed in section 7.4 and the final remarks are presented in section 7.5.

7.2 Basics of the difference-in-differences method

The DID method compares the change in property values with time (T) before (T=0) and after (T=1) the metro started operations for a treated property group (D =1) in comparison to a control property group (D=0). A base DID model is:

$$\ln y_{it} = \alpha + \alpha_{\mathrm{T}}\mathrm{T} + \alpha_D D + \beta(\mathrm{T.D}) + \varepsilon_{it} \quad (6)$$

where:

y_{it} is the value of observation i in time t

T is a dummy time variable that equals 1 if treatment occurred, 0 otherwise

D is a group dummy variable that equals 1 for a treated property, 0 otherwise

α_T represents the effect of time on the changes in the value y

α_D represents the effect of the group (treated or control) on the changes in the value y

β is the DID estimator

The base DID model indicates that the equation reduces to the following for the treated group before and after the treatment, respectively:

$$ln\mathrm{y}_{\mathrm{i}0}^{1} = \alpha + \alpha_1 + \varepsilon_{\mathrm{i}0}^{1} \quad (7)$$

$$\ln \mathrm{y}_{\mathrm{i}1}^{1} = \alpha + \alpha_{\mathrm{T}} + \alpha_1 + \beta + \varepsilon_{\mathrm{i}1}^{1} \quad (8)$$

Therefore, the difference in the values of the treated group before and after the treatment is:

$$(\ln \mathrm{y}_{\mathrm{i}1}^{1} - \ln \mathrm{y}_{\mathrm{i}0}^{1}) = \alpha_{\mathrm{T}} + \beta + (\varepsilon_{\mathrm{i}1}^{1} - \varepsilon_{\mathrm{i}0}^{1}) \quad (9)$$

In the same manner, the values for the control group before and after the treatment are, respectively:

$$\ln y_{i0}^{0} = \alpha + \varepsilon_{i0}^{0} \tag{10}$$

$$\ln y_{i1}^{0} = \alpha + \alpha_{T} + \varepsilon_{i1}^{0} \tag{11}$$

The difference in values will then be:

$$(\ln y_{i1}^{0} - \ln y_{i0}^{0}) = \alpha_{T} + (\varepsilon_{i1}^{0} - \varepsilon_{i0}^{0}) \tag{12}$$

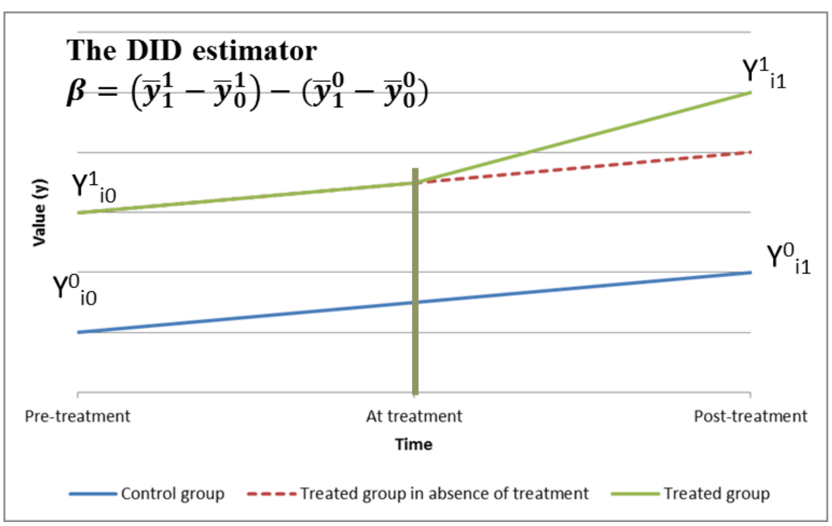
Differencing the differences in values for the treated and control groups removes the effect of group and time and implies the following:

$$(\ln y_{i1}^{1} - \ln y_{i0}^{1}) - (\ln y_{i1}^{0} - \ln y_{i0}^{0}) = \beta + (\varepsilon_{i1}^{1} - \varepsilon_{i0}^{1}) - (\varepsilon_{i1}^{0} - \varepsilon_{i0}^{0}) \tag{13}$$

Therefore, β is the unbiased difference-in-differences (DID) estimator if:

$$E\left[(\varepsilon_{i1}^{1} - \varepsilon_{i0}^{1}) - (\varepsilon_{i1}^{0} - \varepsilon_{i0}^{0})\right] = 0. \tag{14}$$

Figure 7-1: Illustration of the DID concept (source: self-produced graph)



Three assumptions underlie the DID method (e.g. Bertrand et al., 2004; Cameron and Trivedi, 2005). First, the same treated and control groups have to repeat over time in order to allow for a correct DID estimate. As explained in chapters 5 and 6, in this study the same property groups are used before and after the treatment. Second, it is assumed that the trend of the average property value in the treated and control groups would be parallel in the absence of the treatment. Figure 11-5 in appendix E confirms that this assumption is met for the RERA datasets, but it is not possible to check this directly for the REIDIN datasets since data is only available from 2009 (i.e. only one year before operations). Third, and similar to other regression models, the error term in the DID model is assumed to be identically and independently distributed.

7.3 Development of the difference-in-differences models

The structure of the DID model differs across case studies. For example, Agostini and Palmucci (2008) estimate the effect of the metro in Santiago on property values at two stages (at announcement and during construction) using two separate models for each stage. On the other hand, Concas (2012) uses one model to study the effect of a highway project on the value of treated and control properties at different stages of the transport development (i.e. during construction, in the opening year and after the opening year compared to the pre-construction period), by using a dummy variable for each stage.

Wu (2012) also uses one DID model to estimate the effect of accessibility to opened and planned railway stations on land values, and differentiates the impact within four distance bands (0.5 km, 1 km, 2 km and 4 km). In a slightly different model structure, Gibbons and Machin (2005) estimate the effect of distance reduction to a railway station due to the opening of new lines on the value of treated properties located within and beyond 2 km, compared to control properties, but without estimating the effect at various project stages.

The base DID model presented in the previous section assumes no time-varying variables are added to the model. Since genuine panel data was not obtained in this study and since the aim was to test for the effect of the metro using two data structures (repeated cross-sectional data and pseudo panel data), the base DID model is improved here to control for time-varying variables. The reader is referred to Table 6-4 for the full list of variables in each dataset.

As explained in chapter 5, the observations in the datasets are grouped either in clusters of properties (in the repeated cross-sectional data) or cohorts based on property location and size (in the pseudo panel data). Although the same clusters and groups repeat over time, the properties within a group can change. This implies that although building characteristics (e.g. the availability of a gym or a swimming pool) are constant over the study period, property attributes (denoted as the vector X) within a group can change over time. In addition, time-varying location attributes (such as distance to the nearest highway, denoted as L) and neighbourhood attributes (such as employment density, denoted as C) need to be controlled for.

To account for the change in property values in a treated versus a control group for reasons other than accessibility to a metro station, a group specific effect (D) is included in the DID model. In the pseudo panel data, however, we have controlled for the location-specific effect by grouping observations to the same pseudo panel groups that repeat over time. To avoid model over-specification, therefore, the group effect (D) is eliminated for the models using pseudo panel data. Comparing this approach with other studies, Gibbons and Machin (2005) have also grouped the repeated cross-sectional property data to postcode units and, by specifying the fixed effect to the postcode unit, they do not include the group effect as a stand-alone variable in the DID model. In their case, therefore, the group effect (D) only interacts with the time effect (T) before and after the treatment. On the other hand, studies that use every property observation in the repeated cross-sectional data, do include the D variable in the DID model (e.g. Agostini and Palmucci, 2008; Concas, 2012).

For each data structure, two sets of DID models are tested, as shown below. First, the basic DID model examines the effect before and after the opening of the metro using time and group dummies as well as property characteristics. This allows one to consider only the impact of the metro and the physical attributes of the property, without considering the effect of time-varying external factors. The first version of the DID model is enhanced by adding time-varying location attributes as well as controlling for neighbourhood variables, like employment density. The second version is called the full DID model. It is worth mentioning that since DID eliminates the effect of constant variables over time, these variables are not included in the DID models.

Basic DID model using repeated cross-sectional data:

$$lny_{i(j)t} = \alpha + \alpha_T T + \gamma_D D + \beta(T.D) + \alpha_x X_{i(j)t} + \mu_j + \varepsilon_{i(j)t} \quad (15)$$

Basic DID model using pseudo panel data:

$$lny_{jt} = \alpha + \alpha_T T + \beta(T.D) + \alpha_x X_{jt} + \mu_j + \varepsilon_{jt} \quad (16)$$

Full DID model using repeated cross-sectional data:

$$lny_{i(j)t} = \alpha + \alpha_T T + \gamma_D D + \beta(T.D) + \alpha_x X_{i(j)t} + \alpha_L L_{i(j)t} + \alpha_C C_{i(j)t} + \mu_j + \varepsilon_{i(j)t} \quad (17)$$

Full DID model using pseudo panel data:

$$lny_{jt} = \alpha + \alpha_T T + \beta(T.D) + \alpha_x X_{jt} + \alpha_L L_{jt} + \alpha_C C_{jt} + \mu_j + \varepsilon_{jt} \quad (18)$$

Where:

$y_{it,\,j}$ is the value of property i located in building j in time t

y_{jt} is the value of pseudo panel group j in time t

$X_{it,j}$ is a vector of property characteristics for property i located in building j in time t

X_{jt} is a vector of mean values of property characteristics for pseudo panel group j in time t

T takes the value of one if the transaction or listing record is after the opening of the metro, and zero other wise

$L_{it,j}$ is a vector of time-varying location attributes for property i located in building j (this is the distance to the nearest highway)

L_{jt} is a vector of time-varying location attributes for pseudo panel group j (same as above)

$C_{it,j}$ is a vector of the time-varying neighbourhood attributes for property i located in building j (this is employment density)

C_{jt} is a vector of time-varying neighbourhood attributes for pseudo panel group j (same as above)

μ_j controls for time-invariant unobserved heterogeneity related to the building j (repeated cross-sectional data) and the pseudo panel group j (pseudo panel data)

7.4 Results

The results from the basic and full DID models are presented in this section. The findings are discussed separately using each dataset and for the three defined catchment areas: 0.5 km, 1 km and 1.5 km from a metro station. As explained in the previous chapter, the findings from models using the RE estimator (the preferred estimator) are discussed here, although the key results using an FE estimator are also presented in appendix F.

It is worth mentioning that since the DID compares the values before and after the opening of the metro, the findings reveal the average effect over a given period of time. For RERA datasets, property records were obtained for three and two years, before and after the opening of the metro, respectively, hence the results show the average effect of the metro over 2.5 years (1.5 years before and 1 year after the metro). For REIDIN datasets, meanwhile, since records were obtained for one and two years, before and after the metro, respectively, the results reveal the average effect over 1.5 years. The annualized effect of the metro over the study time period is also estimated, therefore, report values in Appendix E and discuss them in relation to the existing literature in section 10.3.

7.4.1 Sale transactions of residential properties

Starting with the effect of the metro on the sale transaction values of residential properties (Table 7-1), the findings suggest comparable coefficient values using the repeated cross-sectional data and pseudo panel data, although the significance level is not always similar. In general, a significant positive impact of the Dubai Metro on the value of residential properties located between 0.5 km and 1.5 km of a metro station was observed.

While the basic DID model using repeated cross-sectional data indicates a negative effect of the metro on the value of properties located within 0.5 km of a station of -5.7%, the null hypothesis of no-effect cannot be rejected for the model using pseudo panel data. Nevertheless, the full model reveals that dwellings situated within 0.5 km experience a reduction in value of 9% and 17.7% using the repeated cross-sectional and pseudo panel data, respectively. Comparing this with the results from the existing empirical work, some other case studies also indicate that railways have a negative effect on the value of residential properties in close proximity to stations (Du and Mulley, 2006; Dubé et al., 2013).

An explanation for the negative impact of the Dubai Metro is the presence of negative externalities due to the increased levels of traffic, noise and pollution in this catchment area. These negative externalities have probably increased due to the additional transport measures that RTA has implemented within 0.5 km of a metro station, such as increasing the number of feeder buses and taxis connecting to the metro.

Turning to the model that defines the metro catchment area as a 1 km radius, and using repeated cross-sectional data, the results suggest a positive effect of the metro on the value of treated residential properties of 10.8% and 7.8% for the basic and full models, respectively. Although the coefficient values for the effect of the metro in the basic and full models using pseudo panel data are positive, the results are not significant. In addition, the findings from the basic model indicate that dwellings located within 1.5 km of a station experience an increase in value of 10% and 8% using the repeated cross-sectional data and pseudo panel data, respectively. Nevertheless using the full models, the null hypothesis of no-effect cannot be rejected regarding the value of properties located within 1.5 km of a metro station.

An explanation for the lack of significance in the results generated using the pseudo panel data is probably the measurement error in the pseudo panel data. As a result of creating cohorts that consist of an average value for observations, some information is lost. The reasons for this and the implications regarding the results are discussed further in chapter 10, since they are common across all the empirical models presented in chapters 7 to 9. In addition, it can be observed that the effect of the metro is reduced in the full model. This is to be expected since using a more comprehensive set of time-varying variables controls for the effect of other factors on the value of properties.

The effects of other variables on property values are also significant in the majority of models. The findings for the group effect (treated versus control) suggest that the value of properties in the treated group is higher by 34% to 60% due to reasons other than accessibility to the metro; such as being in a community that contains better building and infrastructure quality. The findings also indicate that an increase in property area of one square metre uplifts its value by between 0.2% to 0.6%. The results also indicate that a one kilometre reduction in distance to a highway reduces values by 2%. This can be related to the increased levels of noise and pollution associated with proximity to a highway. In addition, it is found that an increase in the level of employment density increases the value of dwellings. This suggests that commercial areas are valued more in Dubai.

Finally, examining the goodness of fit, the full model using repeated cross-sectional data explains at least 34% of the variation in property sale data whereas the full model using pseudo panel data explains at least 23%. In addition, we find that the goodness of fit is higher in the full models compared to the basic models.

7.4.2 Sale listings of residential properties

Moving to the results for the sale listings of residential properties (Table 7-2), the estimates using repeated cross-sectional data and pseudo panel data are comparable in value although not always similar in significance level. While the results reveal a negative effect of the metro on the asking sale value of residential properties located within 0.5 km of a station using repeated cross-sectional data (to a maximum of -1.4%), no noticeable effect is found in the model using pseudo panel data.

The models considering wider catchment areas, however, show a positive effect of proximity to a metro station. The basic and full models using repeated cross-sectional data suggest that the sale listing value of dwellings located within 1 km of a station increases by 3.1% and 2.6%, respectively, whereas only the basic model using pseudo panel data suggests a significant uplift of 6.9%. The results for the 1.5 km catchment area also reveal that dwellings experience an enhancement in value of 2.6% and 8.9% using the full model for the repeated cross-sectional data and pseudo panel data, respectively.

Other covariates also have a significant effect on the value of residential properties. The results for the group effect using the full models suggest that for reasons other than access to a station, such as the improved air quality conditions and building quality, the value of properties located within 1 km is higher than the value of properties located further away, in the range of 11% to 17%. The findings also reveal that the sale listing value of residential properties reduces with time by a minimum of 1.3% using repeated cross-sectional data and by a minimum of 6.5% using pseudo panel data. This is due to the global economic downturn which occurred in the year 2009, which resulted in a general reduction in the sale listing value of dwellings across Dubai (Dubai Municipality, 2010).

In addition, it is found that an increase in the property area by one square metre uplifts the sale asking values by 0.1% and 0.4% for the models using repeated cross-sectional data and pseudo panel data, respectively. The results also show that for each year that a property's age decreases its value increases by 10% and 8% using cross-sectional data and panel data, respectively. Similar to the results for the sale transactions of dwellings, the findings for the sale listings of dwellings indicate that property values increase with an increase in employment density.

Finally, the full DID models' goodness of fit ranges from 22% to 27% for the models using repeated cross-sectional data and from 61% to 63% for the models using pseudo panel data. Although this can imply that the model using pseudo panel data appears to better represent the variation in the sale listing values of residential properties, it will be argued in chapter 10 that the findings using the repeated cross-sectional are in fact more consistent.

7.4.3 Rental listings of residential properties

Turning to the effect of the metro on rent listing values (Table 7-3), the estimates from the basic and full models and the two data structures are quite different. While the impact of the metro decreases with distance in the model using repeated cross-sectional data, the reverse is true for the model using pseudo panel data. The large difference between the estimates using the two data structures is largely related to the measurement error when creating the pseudo panel dataset. Figure 6-8 shows that the larger the distance band, the greater the difference between the average rental values in the two data structures.

The findings using repeated cross-sectional data indicate that properties located within 0.5 km of a station experience an increase in value of about 5% in the basic and full models, whereas the null hypothesis of no-effect cannot be rejected for the full model using pseudo panel data. The estimates using repeated cross-sectional data for rented properties located within 1 km of a metro station suggest a positive, but smaller, effect of the metro on rental values, of 2.6% and 1% using the basic and full models, respectively. The models using pseudo panel data, however, reveal a much larger effect of proximity to a station; the results for basic and full models show that the rental value of dwellings increase by 25% and 8.7%, respectively.

While the results for the 1.5 km distance band using repeated cross-sectional data in the basic model indicate that the metro increases rental values by 2.3%, no noticeable effect is found on the value of properties using the full model. Nevertheless, the results for the pseudo panel data indicate a much larger effect of 35% and 11.6% for the basic and full models, respectively. The findings from the full model might be considered as being more consistent since these models also consider the effect of other time-varying variables.

With regards to the effect of other variables on the rent listing value of dwellings, the results for the repeated cross-sectional and pseudo panel data are generally in line, however, the findings are not always similar using the basic and full models. While the findings for the group effect, which estimates the effect of the unobserved internal heterogeneity of treated versus control groups, indicate that the rental value of treated properties is higher than the value of control properties using the basic model, the opposite is true using the full model. In addition, the results suggest that rental values in Dubai have reduced with time (the years 2010 and 2011 vs. 2009) by a minimum of 2.7% and 9% using repeated cross-sectional data and pseudo panel data, respectively. This is closely related to the economic downturn which has also affected the rental listing value of residential properties across the emirate.

The results also show that an increase in property area of one square metre increases rent values by between 0.2% and 0.5% and that the older the property the lower the rent value by about 3.5% for each additional year. In addition, the results show that the higher the employment density, the higher the rental values of residential properties by between 24% and 38%. While the full models using repeated cross-sectional data explain 45% to 48% of the variation in property rent values, the full models using pseudo panel data explain about 87% of the variation.

7.4.4 Sale transactions of retail properties

In line with the meta-analysis results and as indicated in the literature, the impact of the Dubai Metro is higher on the value of retail properties than on residential properties. The estimates in Table 7-4 show a positive effect across all catchment areas.

The findings using repeated cross-sectional data in the basic and full models suggest that the value of retail properties located within 0.5 km of a station is enhanced by 17.1% and 39.4% respectively, whereas no significant impact is observed in the models using pseudo panel data. An explanation for the lack of significance in the results from the pseudo panel data is the measurement error in averaging retail property observations, as introduced in chapter 5 and as will be explained further in chapter 10.

Moving to the results for wider catchment areas, it can be observed that the value of retail properties located within 1 km and 1.5 km of a metro station is increased by 40.5% to 42%, respectively, using repeated cross-sectional data. As for the model using pseudo panel data, the findings suggest an increase in sale values in the range of 27% to 30.3% for retail properties within 1 km of a station and by 44% to 46% for properties located within 1.5 km.

The results also show significant effects of other variables. Using the repeated cross-sectional data, the coefficient values for the group effect suggest that the value of treated retail properties is higher (by a minimum of about 25%) compared to properties in the control group. This difference is probably due to the unobserved heterogeneity within the treated group. In addition, it is found that sale values are reduced over time by between 11% and 46%, depending on the model type and data structure. This is not surprising, since the global economic recession has also affected the sales value of retail properties in Dubai.

As expected, the effect of the size of a commercial property on its value is higher than that for residential properties; the findings suggest an increase in the value of retail properties with a unit increase in area in a range of 0.4% to 0.6%. In contrast to the results for the sale of residential properties, the findings for retail properties reveal that a one kilometre distance reduction to a highway enhances property values by between 3% and 8% using repeated cross-sectional data. However, the findings show no noticeable effect using pseudo panel data.

The results for the impact of employment density are mixed using repeated cross-sectional data, whereas no significant effect is found from the models using pseudo panel data. Finally, the models explain from 51% to 61% of the variation in the data using repeated cross-sectional data and from 51% to 54% using pseudo panel data.

Table 7-1: Results of the DID models for the effect of the metro on sale transactions of residential properties (RERA)

Catchment zone	**0.5 km**				**1 km**				**1.5 km**			
Data structure	**RCS**		**PP**		**RCS**		**PP**		**RCS**		**PP**	
No. Observations	**39,308**		**3,344**		**39,308**		**3,344**		**39,308**		**3,344**	
Model type	**Basic**	**Full**	**Basic**	**Full**	**Basic**	**Full**	**Basic**	**Full**	**Basic**	**Full**	**Basic**	**Full**
Observation years	2007-2011											
Covariates												
Constant	12.660*** (0.021)	11.960*** (0.065)	13.428*** (0.034)	13.014*** (0.090)	12.656*** (0.021)	12.167*** (0.066)	13.419*** (0.034)	13.033*** (0.094)	12.697*** (0.021)	12.264*** (0.067)	13.406*** (0.033)	13.061*** (0.095)
Group effect	NA	0.460*** (0.055)	NA	NA	NA	0.335*** (0.038)	NA	NA	NA	0.599*** (0.033)	NA	NA
Time effect	-0.015*** (0.004)	-0.015*** (0.005)	0.041*** (0.015)	0.077*** (0.021)	-0.044*** (0.005)	-0.035*** (0.0006)	0.029* (0.016)	0.065*** (0.025)	-0.088*** (0.008)	-0.036*** (0.011)	0.015 (0.017)	0.048* (0.027)
Group * time effect (DID)	-0.057*** (0.014)	-0.090*** (0.014)	-0.036 (0.051)	-0.177* (0.054)	0.108*** (0.007)	0.078*** (0.008)	0.055 (0.034)	0.043 (0.039)	0.101*** (0.010)	0.018 (0.013)	0.084*** (0.030)	0.033 (0.038)
Area (sq.m)	0.006*** (0.00007)	0.006*** (0.00007)	0.002*** (0.0001)	0.003*** (0.0001)	0.006*** (0.00007)	0.006*** (0.00007)	0.002*** (0.0001)	0.002*** (0.0001)	0.006*** (0.00007)	0.006*** (0.00007)	0.002*** (0.0001)	0.002*** (0.0001)
Distance to the nearest highway (km)	NA	0.02*** (0. 001)	NA	0.01*** (0.004)	NA	0.01*** (0.001)	NA	0.01*** (0.001)	NA	0.02*** (0.001)	NA	0.01* (0.005)
Log employment density	NA	0.103*** (0.011)	NA	0.056*** (0.012)	NA	0.066*** (0.011)	NA	0.054*** (0.012)	NA	0.023** (0.011)	NA	0.050*** (0.013)
R^2 *within*	0.1632	0.1669	0.0051	0.0048	0.1655	0.1671	0.0034	0.0038	0.1631	0.1654	0.0016	0.0027
R^2 *between*	0.2590	0.2713	0.2427	0.2621	0.2674	0.2939	0.2506	0.2636	0.2699	0.3470	0.2633	0.2687
R^2 *overall*	0.3297	0.3361	0.2079	0.2872	0.3388	0.3467	0.2133	0.2366	0.3423	0.3855	0.2214	0.2275

Legend: RCS: repeated cross-sectional data. PP: Pseudo panel data

*, **, *** indicate significance at 10%, 5% and 1% respectively. Values in parentheses are standard errors.

Table 7-2: Results of the DID models for the effect of the metro on sale listings of residential properties (REIDIN)

Catchment zone	**0.5 km**				**1 km**				**1.5 km**			
Data structure	**RCS**		**PP**		**RCS**		**PP**		**RCS**		**PP**	
No. Observations	**165,978**		**2,288**		**165,978**		**2,288**		**165,978**		**2,288**	
Model type	**Basic**	**Full**	**Basic**	**Full**	**Basic**	**Full**	**Basic**	**Full**	**Basic**	**Full**	**Basic**	**Full**
Observation years	2009-2011											
Covariates												
Constant	12.956*** (0.017)	12.387*** (0.048)	12.569*** (0.025)	11.862*** (0.053)	12.914*** (0.023)	12.349*** (0.051)	12.555*** (0.025)	11.871*** (0.054)	12.698*** (0.026)	12.418*** (0.049)	12.515*** (0.023)	11.913*** (0.053)
Group effect	0.221*** (0.039)	0.111*** (0.038)	NA	NA	0.139*** (0.030)	-0.167*** (0.038)	NA	NA	0.395*** (0.030)	0.008 (0.046)	NA	NA
Time effect	0.004 (0.003)	-0.013*** (0.003)	-0.073*** (0.013)	-0.065*** (0.013)	-0.020*** (0.004)	-0.038*** (0.005)	-0.113*** (0.016)	-0.073*** (0.016)	-0.033*** (0.005)	-0.034*** (0.005)	-0.200*** (0.018)	-0.133*** (0.018)
Group * time effect (DID)	-0.013** (0.006)	-0.014** (0.006)	0.026 (0.021)	0.009 (0.021)	0.031*** (0.004)	0.026*** (0.005)	0.069*** (0.017)	0.017 (0.017)	0.046*** (0.005)	0.026*** (0.004)	0.160*** (0.019)	0.089*** (0.019)
Area (sq.m)	0.001*** (0.00001)	0.001*** (0.00001)	0.004*** (0.0001)	0.004*** (0.0001)	0.001*** (0.00001)	0.001*** (0.00001)	0.004*** (0.0001)	0.004*** (0.00009)	0.001*** (0.00001)	0.001*** (0.00001)	0.004*** (0.0001)	0.004*** (0.0001)
Age	-0.096*** (0.002)	-0.097*** (0.002)	-0.057*** (0.006)	-0.077*** (0.006)	-0.096*** (0.002)	-0.097*** (0.002)	-0.056*** (0.006)	-0.077*** (0.006)	-0.096*** (0.002)	-0.096*** (0.002)	-0.049*** (0.006)	-0.073*** (0.006)
Log employment density	NA	0.101*** (0.008)	NA	0.123*** (0.009)	NA	0.128*** (0.010)	NA	0.122*** (0.009)	NA	0.097*** (0.011)	NA	0.110*** (0.009)
R^2 *within*	0.1249	0.1248	0.3490	0.3646	0.1252	0.1249	0.3454	0.3649	0.1254	0.1249	0.3356	0.3619
R^2 *between*	0.1696	0.2778	0.5585	0.6380	0.1705	0.2809	0.5749	0.6380	0.2537	0.2730	0.6244	0.6545
R^2 *overall*	0.2293	0.2739	0.5431	0.6161	0.2296	0.2588	0.5571	0.6160	0.1392	0.2228	0.5983	0.6300

Legend: RCS: repeated cross-sectional data. PP: Pseudo panel data

*, **, *** indicate significance at 10%, 5% and 1% respectively. Values in parentheses are standard errors.

Table 7-3: Results of the DID models for the effect of the metro on rental listings of residential properties (REIDIN)

Catchment zone	0.5 km				1 km				1.5 km			
Data structure	**RCS**		**PP**		**RCS**		**PP**		**RCS**		**PP**	
No. Observations	**81,248**		**823**		**81,248**		**823**		**81,248**		**823**	
Model type	**Basic**	**Full**	**Basic**	**Full**	**Basic**	**Full**	**Basic**	**Full**	**Basic**	**Full**	**Basic**	**Full**
Observation years	2009-2011											
Covariates												
Constant	7.444*** (0.023)	5.574*** (0.055)	6.940*** (0.041)	5.520*** (0.050)	7.207*** (0.025)	5.416*** (0.064)	6.888*** (0.035)	5.595*** (0.055)	7.110*** (0.023)	13.591*** (0.759)	6.867*** (0.032)	5.637*** (0.056)
Group effect	0.296*** (0.053)	-0.062* (0.036)	NA	NA	0.544*** (0.034)	-0.166*** (0.034)	NA	NA	0.636*** (0.030)	-0.140*** (0.046)	NA	NA
Time effect	-0.072*** (0.005)	-0.130*** (0.005)	-0.122*** (0.021)	-0.091*** (0.017)	-0.088*** (0.006)	-0.130*** (0.006)	-0.255*** (0.024)	-0.132*** (0.020)	-0.090*** (0.006)	-0.027** (0.120)	-0.336*** (0.025)	-0.155*** (0.022)
Group * time effect (DID)	0.051*** (0.013)	0.051*** (0.013)	0.081** (0.032)	0.037 (0.027)	0.026*** (0.006)	0.010* (0.006)	0.249*** (0.026)	0.087*** (0.023)	0.025*** (0.006)	0.008 (0.006)	0.351*** (0.027)	0.116*** (0.025)
Area (sq.m)	0.002*** (0.00002)	0.002*** (0.00002)	0.005*** (0.0002)	0.005*** (0.0002)	0.002*** (0.00002)	0.002*** (0.00002)	0.005*** (0.0002)	0.005*** (0.0002)	0.002*** (0.00002)	0.002*** (0.00002)	0.005*** (0.0002)	0.005*** (0.0002)
Age	-0.034*** (0.002)	-0.034*** (0.002)	0.010 (0.010)	-0.036*** (0.007)	-0.030*** (0.002)	-0.035*** (0.002)	0.021** (0.009)	-0.033*** (0.007)	-0.029*** (0.002)	-0.033*** (0.002)	0.022** (0.008)	-0.032*** (0.006)
Log employment density	NA	0.328*** (0.009)	NA	0.265*** (0.009)	NA	0.338*** (0.012)	NA	0.251*** (0.010)	NA	0.379*** (0.015)	NA	0.242*** (0.010)
R^2 *within*	0.1039	0.1038	0.1879	0.2197	0.1038	0.1036	0.1878	0.2327	0.1037	0.1035	0.1789	0.2364
R^2 *between*	0.2916	0.7329	0.6258	0.9069	0.4236	0.7419	0.7270	0.9080	0.4908	0.7508	0.7816	0.9100
R^2 *overall*	0.3035	0.4511	0.5925	0.8649	0.1608	0.4683	0.6837	0.8668	0.1282	0.4818	0.7287	0.8685

Legend: RCS: repeated cross-sectional data. PP: Pseudo panel data

*, **, *** indicate significance at 10%, 5% and 1% respectively. Values in parentheses are standard errors.

Table 7-4: Results of the DID models for the effect of the metro on sale transactions of retail properties (RERA)

Catchment zone	0.5 km				1 km				1.5 km			
Data structure	**RCS**		**PP**		**RCS**		**PP**		**RCS**		**PP**	
No. Observations	**3,419**		**336**		**3,419**		**336**		**3,419**		**336**	
Model type	**Basic**	**Full**	**Basic**	**Full**	**Basic**	**Full**	**Basic**	**Full**	**Basic**	**Full**	**Basic**	**Full**
Observation years	2007-2011											
Covariates												
Constant	12.932*** (0.076)	12.709*** (0.210)	13.175*** (0.074)	13.022*** (0.196)	12.938*** (0.078)	13.068*** (0.195)	13.191*** (0.070)	13.128*** (0.190)	12.744*** (0.127)	13.307*** (0.200)	13.196*** (0.070)	13.282*** (0.195)
Group effect	0.448*** (0.114)	0.444*** (0.116)	NA	NA	0.245** (0.099)	0.280*** (0.100)	NA	NA	0.312** (0.138)	0.756*** (0.176)	NA	NA
Time effect	-0.300*** (0.020)	-0.393*** (0.025)	-0.109** (0.043)	-0.138** (0.064)	-0.417*** (0.023)	-0.448*** (0.025)	-0.222*** (0.053)	-0.263*** (0.067)	-0.462*** (0.024)	-0.443*** (0.026)	-0.450*** (0.085)	-0.448*** (0.090)
Group * time effect (DID)	0.171*** (0.046)	0.394*** (0.054)	0.115 (0.083)	0.166 (0.114)	0.406*** (0.037)	0.420*** (0.037)	0.273*** (0.071)	0.303*** (0.078)	0.417*** (0.036)	0.405*** (0.036)	0.444*** (0.092)	0.458*** (0.096)
Area (sq.m)	0.006*** (0.0004)	0.006*** (0.0004)	0.005*** (0.0004)	0.005*** (0.0004)	0.006*** (0.0004)	0.006*** (0.0004)	0.004*** (0.0004)	0.004*** (0.0004)	0.006*** (0.0004)	0.006*** (0.0004)	0.004*** (0.0004)	0.005*** (0.0004)
Distance to the nearest highway (km)	NA	-0.08*** (0.01)	NA	-0.01 (0.01)	NA	-0.05*** (0.01)	NA	-0.001 (0.02)	NA	-0.03** (0.01)	NA	0.01 (0.02)
Log employment density	NA	0.102*** (0.035)	NA	0.036 (0.038)	NA	0.012 (0.033)	NA	0.028 (0.033)	NA	-0.141*** (0.042)	NA	-0.023 (0.034)
R^2 *within*	0.1031	0.1167	0.0570	0.0530	0.1361	0.1288	0.0621	0.0683	0.1338	0.1377	0.1005	0.0980
R^2 *between*	0.6097	0.5963	0.5635	0.5692	0.6054	0.6191	0.6006	0.5994	0.5902	0.6105	0.5968	0.6010
R^2 *overall*	0.5147	0.5080	0.5012	0.5070	0.5699	0.6065	0.5389	0.5379	0.6171	0.5366	0.5402	0.5441

Legend: RCS: repeated cross-sectional data. PP: Pseudo panel data

*, **, *** indicate significance at 10%, 5% and 1% respectively. Values in parentheses are standard errors.

7.5 Conclusions

This chapter has presented the estimates for the models testing for the effect of proximity to a metro station via a DID estimator. The DID does not only estimate the relationship between the value of properties and the various factors that can affect it, but it also controls for potential causality between regressors and the dependent variable. Two types of DID model are tested: the first is a basic model that only controls for group and time effects as well as property characteristics. The second also includes all time-varying variables to control for the effect of the measured and available heterogeneity between treated and control groups on the value of properties. Since the full DID models include more regressors, the results are also more realistic and robust.

It can be observed from the results summarised above that only a small number of the findings from the repeated cross-sectional data and the pseudo panel data are similar. Although the pseudo panel data was created in such a way as to maintain a level of homogeneity within a cohort and increase the variation across cohorts, the datasets only allowed a limited cohort size (less than 50 in the majority of the RERA datasets and less than 100 in the majority of the REIDIN datasets). As a result, the loss of information on individual records in the pseudo panel dataset has led to measurement error and biased the results. The repeated cross-sectional data, on the other hand, controls for the unobserved heterogeneity and produces more reasonable estimates for the effect of the Dubai Metro on the value of residential and retail properties. These are discussed further in chapter 10.

The findings from the preferred models suggest that the effect of the metro on the sale transaction value and sale listing value of residential properties located within 0.5 km of a metro station is negative, whereas it is positive for properties located within 1 km and 1.5 km from a station. On the other hand, the estimates are positive for the rental value of dwellings located within a 0.5 km and 1 km distance from a metro station, but insignificant for properties in the 1.5 km distance band. Finally, it was found that the metro has enhanced the sale transaction value of treated retail properties located at all catchment areas around a station.

The results from the DID models imply that while negative externalities such as the increase in the levels of noise, pollution and traffic at close distances to a metro station can outweigh the accessibility benefit to the metro for the sale value of dwellings, the opposite is true for the rental value of dwellings and for the sale value of retail properties. The next chapter estimates the effect of proximity to one or more metro stations on property values using the conventional method of hedonic pricing that is corrected for sources of potential bias.

Chapter 8. ESTIMATING THE EFFECT OF PROXIMITY TO ONE OR MORE METRO STATIONS

8.1 Introduction

The hedonic pricing (HP) method is the most widely used method to estimate the effect of a transport system on land and property values. In this chapter, we test for the effect of accessibility to one or more metro stations by applying HP models. Similar to the previous chapter, the effect of the metro is examined using the repeated cross-sectional data and pseudo panel data from each of the four selected property datasets.

The effect of the metro is estimated on the value of properties within the main catchment areas of 0.5 km, 1 km and 1.5 km. Since sufficient property and transport data is available for a more refined HP model, albeit only for models using repeated cross-sectional data, the effect of proximity at smaller catchment areas is also estimated so as to observe the variation in results with distance to a station. The findings from this chapter should in theory be aligned with the results from the DID models.

This chapter is structured as follows. Section 8.2 presents the HP models applied in this study. The results generated using each dataset are presented in section 8.3. The final remarks from this chapter are set out in section 8.4.

8.2 Model specification

In this section, HP models are developed to test for the effect of proximity to one or more metro stations, controlling for those variables that can affect property values and which are available for this study (Table 6-4). Repeated cross-sectional data corrected for unobserved heterogeneity across properties as well as pseudo panel data are also used. Two versions of HP models, which differ in the definition of the catchment area around a metro station, are tested.

In the first version, similar treatment catchment areas as were defined for the DID models are used, with a separate model for each catchment area. Hence, the effect of proximity to metro stations on the value of properties located at each distance band is tested. For the first distance band, the model assumes that only the properties located within 0.5 km of a station have access to a metro station, and all other properties do not. In the second and third models, all properties located within 1 km and 1.5 km of a station, respectively, have access to a metro station and the ones located further away do not have access.

A variable N_j is used to define the number of accessible metro stations for observation “j" located within a predefined catchment area. The index “j” refers to the building that contains property “i” in the repeated cross-sectional data and the pseudo panel group in the pseudo panel data. The model specification implies that, for example, a property located within 900 m of one metro station and 950 m from another will have an N_j value of zero for the model using a catchment area of 0.5 km, but a value of 2 for the models using the catchment areas of 1 km and 1.5 km. In other words, some properties that are defined as not accessible in one catchment area become accessible in the models using higher distance bands.

This model specification is chosen for the following main reasons. First, it allows the results from the HP models to be easily compared with the findings from the DID models (which also defines treated properties at the three catchment areas). Second, in testing for the assumption of an accessible catchment area by measuring the effect of proximity to one or more stations on the value of properties located within that distance band, it also then becomes possible to compare the effect across different catchment areas. It may be that consumers bid less for the benefit of accessibility to an additional station for properties located within 1.5 km of a station than for properties located within 1 km.

Property characteristics denoted as vector X are also controlled for in the HP models. In addition, location attributes, denoted as L, are added, controlling for the effect of proximity to amenities (schools, hospitals and shopping areas). To reduce the potential for bias due to between community heterogeneity as well as potential confounding, neighbourhood attributes (C), accounting for employment density and number of shopping trips to a neighbourhood are also included. While the variable L varies across buildings (i.e. properties from different buildings), the variable C varies only across communities. This is the case because the most disaggregated values for the variable C are available only at a community level.

In addition, other unobserved factors related to a cohort, such as the perception of residents regarding use of the metro, are captured within the cohort fixed effect parameter (μ_j) and it is assumed that this remains constant over the study period. This is a reasonable assumption given that the study period is relatively short (5 years maximum). Since property value may also change over time for reasons other than accessibility to transport or amenities, the time effect is controlled for by adding year dummies (t). The equations below present the first version of the HP models for the repeated cross-sectional data and the pseudo panel data.

First version of the HP models using repeated cross-sectional data:

$$lny_{i(j)t} = \alpha + \alpha_t t + \beta N_j + \alpha_x X_{i(j)t} + \alpha_L L_{i(j)t} + \alpha_C C_{i(j)t} + \mu_j + \varepsilon_{i(j)t} \tag{19}$$

First version of the HP models using pseudo panel data:

$$lny_{jt} = \alpha + \alpha_t t + \beta N_j + \alpha_x X_{jt} + \alpha_L L_{jt} + \alpha_C C_{jt} + \mu_j + \varepsilon_{jt} \tag{20}$$

Where:

$y_{i(j)t}$ is the value of property i located in building j in time t

y_{jt} is the value of pseudo panel group j in time t

N_j is the number of accessible metro stations for observation "j" located within a predefined catchment area. The index "j" refers to the building that contains property "i" in the repeated cross-sectional data and the pseudo panel group in the pseudo panel data.

$X_{i(j)t}$ is a vector of property characteristics for property i located in building j in time t

X_{jt} is a vector of mean values of property characteristics for pseudo panel group j in time t

t is the year of the transaction or listing and controls for the variation in property value with time that is common across all locations in Dubai

$L_{i(j)t}$ is a vector of location attributes for property i located in building j in time t (this includes distance to the nearest school, local shops, nearest highway)

L_{jt} is a vector of location attributes for pseudo panel group j in time t (same as above)

$C_{i(j)t}$ is a vector of the neighbourhood attributes for property i located in building j in time t (these are employment and shopping densities)

C_{jt} is a vector of neighbourhood attributes for the pseudo panel group j in time t (same as above)

μ_j controls for time-invariant unobserved heterogeneity related to the building j (repeated cross-sectional data) and the pseudo panel group j (pseudo panel data)

The DID and HP models presented earlier estimate the impact of proximity to a metro station on the value of properties located at pre-defined distance bands for every 500 m radius. This approach is similar to the approach used in the majority of previous empirical work (e.g. Billings, 2011; Bowes and Ihlanfeldt, 2001; Duncan, 2008; Efthymiou and Antoniou, 2013; Pan and Zhang, 2008; Weinberger, 2001).

Although an attempt was made to estimate the effect at smaller catchment areas around a metro station, the limited number of observations in the majority of the pseudo panel datasets has limited the number of catchment zones. This is also the case in most of the other studies. For example, while Agostini and Palmucci (2008) used one catchment area (1 km distance), Billings (2011) divided property data into two catchment areas (within half a mile (0.8 km) and one mile (1.6 km) of a station), and Bowes and Ihlanfeldt (2001) used a range of distance bands from a quarter of a mile (0.4 km) to 3 miles (4.8 km) from a station. On the other hand, the larger number of records in the repeated cross-sectional data makes it possible to estimate the impact of the metro at smaller distance bands using the HP model, by defining dummy variables for each distance band. Setting a threshold of a minimum of 20 observations per catchment area, it was possible to obtain enough property records starting at 300 m from a station and at 200 m consecutive catchments following this. Table 11-6 in appendix D presents the sample size at these smaller catchment areas. The equation below sets out this model, in which a set of dummy variables d are used to identify the distance band in which a property is located. This model is the second version of the HP models.

Second version of the HP model using repeated cross-sectional data:

$$lny_{i(j)t} = \alpha + \alpha_t t + \sum_k \beta d_{jk} + \alpha_x X_{i(j)t} + \alpha_L L_{i(j)t} + \alpha_C C_{i(j)t} + \mu_j + \varepsilon_{i(j)t} \quad (21)$$

8.3 Results

This section presents the results of equations 19 to 21 for each dataset (Table 8-1 to Table 8-4 for results using the first version of the HP models, as well as Table 8-5 and

Figure 8-1 for results using the second version of the HP model). The findings for the effect of accessibility to the metro on the value of properties located within the specified catchment areas are presented separately for each catchment area.

Similar to the DID results, the annualized effect of the metro is also reported. Since the HP models consider the effect over the study time period, the findings are the average effect of the metro over 5 years in RERA datasets and 3 years in REIDIN datasets. The uplift is also reported on an annualized basis in Appendix E.

8.3.1 Sale transactions of residential properties

The results for the effect of proximity to one or more metro stations on the sale value of residential properties using repeated cross-sectional property data are comparable with the results from the DID models. We find that the metro reduces values for properties close to a station, but enhances values for dwellings located between 0.5 km and 1 km (i.e. within walking distances to the station but not too close to it). A smaller positive effect is also estimated for residential properties located within a 1.5 km distance. Nevertheless, no significant estimates were found using pseudo panel data, even though the coefficient values were comparable with those from the models using repeated cross-sectional data.

Starting with the first version of the HP models, a negative effect of the metro is evident on the value of dwellings located within 0.5 km of a station (-8.6%), which is probably due to the increased negative externalities affecting dwellings located within this catchment area. As the accessible catchment area around a metro station is increased, the results suggest a positive effect of accessibility to one or more metro stations of 1.9% and 1.2% for dwellings located within 1 km and 1.5 km, respectively.

Controlling for other variables, it is observed that property values rise significantly due to an increase in property area (a minimum of 0.3% for a one square metre increase), access to schools (a minimum of 20% for a one kilometre distance reduction) and due to an increase in the employment density in the neighbourhood. The values reduce, however, with proximity to shops and hospitals, which can be related to the increase of noise and pollution levels near these amenities. In addition, no significant effect on the value of dwellings is observed due to an increase in the shopping trips to the community where the properties are located. Proximity to a highway, however, reduces the value of residential properties slightly, which is to be expected since the levels of traffic and noise increase near highways. These models explain 38% and 34% of the variation in residential property data using repeated cross-sectional data and pseudo panel data, respectively.

The findings for the second version of the HP model (the more finely delineated distance bands) reveal a concave shape to the relationship between property values and the distance to the metro. The results reveal insignificant estimates for the impact of the metro on properties located within 300 m of a station, between 501-700 m and greater than 1.1 km. The findings, however, suggest a negative impact for dwellings located within a catchment area of 301 m to 500 m (-6%) and a positive and significant effect for dwellings between 701 m and 1.1 km in the range of 10% to 13%, with a peak effect for dwellings within 701 m to 900 m. The concave effect of the metro has also been suggested in few previous studies (e.g. Du and Mulley, 2006; McCann, 2001). This model explains 33% of the variation in property data.

8.3.2 Sale listings of residential properties

Looking at the effect of the metro on the sale listing values of residential properties (

Table 8-2), a very similar impact is found to that for sale transaction values; i.e. a negative effect for properties very close to a station, an increasingly positive effect up to about 1 km from a station and a decreasing positive effect for properties located between 1 and 1.5 km. In addition, there is a difference in the significance power between the two data structures.

Using repeated cross-sectional data, the impact is negative for properties located very close to the metro (-1.4%) but positive for properties situated within 1 km (1.5%) and 1.5 km (0.4%). In contrast, the impact is not significant using pseudo panel data, except for the model defining accessible properties to a station as those located within a 1.5 km catchment area, where a positive effect of 1.9% is found. This lack of statistical significance is probably due to the measurement error in the pseudo panel data, which is discussed more in chapter 10, and which may also have led to the difference in the average property values between the repeated cross-sectional data and pseudo panel data.

Some of the other attributes also substantially affect the asking values of dwellings. The larger the area of the property, the higher its value, in the range of 0.1% to 0.5% per additional square metre. However, no noticeable effect is found regarding the age of the property, except for the models using repeated cross-sectional data on a catchment area of 1 km; where the value of properties increases with age. This can be related to the fact that some properties (about 8% in our datasets) were actually not ready for occupation at the time of initial sale, so when they were resold at or near completion date, the values were higher.

With regards to building attributes, it is found that while the availability of a gym in the building reduces the values of dwellings in general, the availability of a swimming pool increases values, although only substantially in the model using pseudo panel data. In addition, the results indicate that the availability of a porter in the building enhances property values massively by a range of 55% to 87%. It may be that buildings with porter services may also have a much higher finishing quality and may have other unobserved qualities such as a room service or free laundry rooms.

The results also show that proximity to schools increases property values, while proximity to a hospital reduces values (perhaps due to the increased level of noise) and that proximity to shops does not always enhance property values. Finally, the findings indicate that while the increase in employment density does not affect the sale listing value of dwellings, the increase in the number of shopping commuters to the community increases values by a range of 2% to 7.8%. The overall explanatory power of the models ranges between 45% and 72%.

Moving to the results of the second version of the HP model, a positive effect of the metro is found on the value of dwellings located between 301 m and 1.3 km of a station, while the results are insignificant for properties at closer and further distances. The findings indicate a gradual increase in values, reaching a peak at the catchment area from 901 m to 1.1 km, where an increase of almost 7% is estimated. Finally, this model explains about 46% of the variation in property data.

8.3.3 Rental listings of residential properties

Moving to the effect of the metro on the rental listing values of residential properties, the results using repeated cross-sectional data show a positive and reducing effect with distance to one or more metro stations. The findings suggest that the metro increases the rental values of dwellings by 4%, 1.6% and 1% for properties located within 0.5 km, 1 km and 1.5 km of a station, respectively. Nonetheless, the results from the models using pseudo panel data reveal no significant impact of the metro on rental values except for the model defining accessible properties to the metro as those located within 1.5 km, where an increase of 3% is observed. Similar to the results from the DID model, the difference in the estimates between the two data structures is probably due to the loss of information in the pseudo panel data.

Property attributes also affect rental asking values. It is found that an increase in a unit area of properties increases values by 0.2% and 0.5% using repeated cross-sectional and pseudo panel data, respectively. In addition, the null hypothesis of no-effect cannot be rejected for property age, the availability of a gym and a swimming pool. However, the results suggest that the availability of porter services increases rental values by a maximum of 78% and 20% using repeated cross-sectional data and pseudo panel data, respectively. As explained for the sale listing values, this may be related to the other unobserved benefits and quality of the buildings with porter services compared to other properties in Dubai.

Moving to the impact of proximity to amenities, while it is found that proximity to schools and hospitals does not affect the rental value of dwellings, proximity to shops increases values by about 21% using repeated cross-sectional data. The findings also reveal that the greater the employment density, the larger the rental values in the range of 10.3% to 35%. Similarly, in the majority of models, it is found that the rental value of dwellings is higher in the range of 5.4% to 13.5% in communities with a higher number of shopping trips. Finally, the models' explanatory power is 61% and 88% using repeated cross-sectional data and pseudo panel data, respectively.

In contrast to the first version of the HP models (Table 8-3) where the positive impact of the metro on rental values reduces with distance, the second version of the HP model (Table 8-5) indicates that the effect is also positive but peaks at two catchment areas (less than 300 m and from 0.7 km to 0. 9km, where increases in values of 8.6% and 9.5%, respectively, are observed). The results are also positive, although smaller, for dwellings located between 0.3 km and 0.5 km and from 0.9 km to 1.1 km but are not significant at other distances. This model explains 60% of the variation in rent data.

8.3.4 Sale listings of retail properties

Estimating the effect of accessibility to one or more stations on the sale value of retail properties, a positive effect is observed that peaks some distance away from a station. The impact is also much larger on the value of retail properties than it is on the value of residential properties.

Starting with the first version of the HP model on a catchment area of 0.5 km, the results suggest an increase in values of 36% and 23% using the repeated cross-sectional and pseudo panel data, respectively. When defining the accessible properties to a metro station as those located within a 1 km distance, the model estimates indicate that accessibility to an additional station increases values by 34% using repeated cross-sectional data and by 25% using pseudo panel data. The effect of accessibility extends further to retail properties situated within 1.5 km of a station, where the models suggest increases in values from 13.7% to 17.5% depending on the data structure.

As expected, the greater the area of a retail property the higher its value, in the range of 0.4% to 0.6%. The majority of coefficients also indicate that a one kilometre reduction in distance to a highway increases the value of retail properties by 4% to 8%. In addition, it is revealed that the greater the employment density and the number of shopping trips, the lower the value of retail properties. This can be related to an increase in supply, which leads to a reduction in the value of other retail properties in the area. Finally, the models explain between 59% and 63% of the variation in retail property data.

The second version of the HP model finds a positive, significant and concave effect of the metro on the value of retail properties located up to 1.5 km from a station. The results show that the peak effect is for properties located between 701 m to 900 m (+76%) and that the minimum positive impact of the metro on the value of retail properties is 25%. The overall explanatory power of the model is 56%.

Table 8-1: Results of the first version of the HP models for the effect of the metro on sale transactions of residential properties (RERA)

Catchment zone	**0.5 km**		**1 km**		**1.5 km**	
Data structure	**RCS**	**PP**	**RCS**	**PP**	**RCS**	**PP**
No. Observations	**39,308**	**3,344**	**39,308**	**3,344**	**39,308**	**3,344**
Model type	**First version of HP model**	**First version of HP model**	**First version of HP model**	**First version of HP model**	**First version of HP model**	**First version of HP model**
Observation years	2007 - 2011					
Covariates						
Constant	11.882*** (0.119)	13.031*** (0.145)	11.949*** (0.120)	13.046*** (0.147)	11.962*** (0.119)	13.044*** (0.147)
Year 2008	0.313*** (0.009)	0.479*** (0.025)	0.318*** (0.009)	0.480*** (0.025)	0.318*** (0.009)	0.480*** (0.026)
Year 2009	0.273*** (0.010)	0.388*** (0.027)	0.281*** (0.010)	0.389*** (0.027)	0.282*** (0.010)	0.390*** (0.027)
Year 2010	0.259*** (0.010)	0.359*** (0.029)	0.237*** (0.012)	0.353*** (0.031)	0.228*** (0.012)	0.353*** (0.032)
Year 2011	0.227*** (0.010)	0.419*** (0.029)	0.205*** (0.011)	0.413*** (0.031)	0.196*** (0.014)	0.413*** (0.033)
Accessibility to a number of metro stations	-0.086*** (0.013)	-0.031 (0.051)	0.019* (0.010)	0.004 (0.034)	0.012** (0.005)	0.002 (0.015)
Area (sq.m)	0.006*** (0.00006)	0.003*** (0.0001)	0.006*** (0.00006)	0.003*** (0.0001)	0.006*** (0.00006)	0.003*** (0.0001)
Distance to the nearest school (km)	-0.3*** (0.02)	-0.2*** (0.02)	0.3*** (0.02)	-0.2*** (0.02)	-0.3*** (0.02)	-0.2*** (0.02)
Distance to the nearest hospital (km)	0.07*** (0.02)	0.03 (0.02)	0.06*** (0.02)	0.03 (0.02)	0.06*** (0.02)	0.03 (0.02)
Distance to the nearest shops (km)	0.09*** (0.02)	0.04* (0.03)	0.09*** (0.02)	0.04* (0.03)	0.09*** (0.02)	0.04* (0.03)
Distance to the nearest highway (km)	0.02*** (0.001)	0.01** (0.003)	0.01*** (0.001)	0.01** (0.003)	0.01*** (0.001)	0.01* (0.005)
Log employment density	0.155*** (0.011)	0.097*** (0.012)	0.151*** (0.014)	0.096*** (0.012)	0.148*** (0.011)	0.096*** (0.012)
Log shopping trips	0.023 (0.014)	-0.033** (0.016)	0.021 (0.014)	-0.033** (0.016)	0.022 (0.014)	-0.033** (0.016)
R^2 *within*	0.1922	0.1145	0.1912	0.1140	0.1911	0.1138
R^2 *between*	0.3582	0.3861	0.3594	0.3866	0.3599	0.3869
R^2 *overall*	0.3766	0.3411	0.3805	0.3413	0.3818	0.3415

Legend: RCS: repeated cross-sectional data. PP: Pseudo panel data

*, **, *** indicate significance at 10%, 5% and 1% respectively. Values in parentheses are standard errors.

Table 8-2: Results of the first version of the HP models for the effect of the metro on sale listings of residential properties (REIDIN)

Catchment zone	**0.5 km**		**1 km**		**1.5 km**	
Data structure	**RCS**	**PP**	**RCS**	**PP**	**RCS**	**PP**
No. Observations	**165,978**	**2,288**	**165,978**	**2,288**	**165,978**	**2,288**
Model type	**First version of HP model**	**First version of HP model**	**First version of HP model**	**First version of HP model**	**First version of HP model**	**First version of HP model**
Observation years	2007 – 2011					
Covariates						
Constant	11.725*** (0.074)	11.516*** (0.073)	11.737*** (0.073)	11.518*** (0.073)	11.734*** (0.073)	11.542*** (0.074)
Year 2010	-0.103*** (0.007)	-0.061*** (0.012)	-0.116*** (0.008)	-0.066*** (0.014)	-0.111*** (0.008)	-0.092*** (0.017)
Year 2011	-0.216*** (0.014)	-0.249*** (0.017)	-0.230*** (0.014)	-0.254*** (0.019)	-0.225*** (0.014)	-0.281*** (0.020)
Accessibility to a number of metro stations	-0.014** (0.006)	-0.011 (0.020)	0.015*** (0.004)	0.005 (0.012)	0.004*** (0.002)	0.019** (0.007)
Area (sq.m)	0.001*** (0.00001)	0.005*** (0.0001)	0.001*** (0.000001)	0.004*** (0.00009)	0.001*** (0.0000001)	0.004*** (0.00009)
Age	0.012* (0.007)	-0.004 (0.007)	0.013*** (0.007)	-0.005 (0.007)	0.013* (0.007)	-0.005 (0.007)
Availability of a gym	-0.090*** (0.028)	-0.076** (0.036)	-0.092*** (0.028)	-0.075** (0.036)	-0.091*** (0.028)	-0.077* (0.036)
Availability of a swimming pool	0.028 (0.021)	0.126*** (0.025)	0.028 (0.021)	0.126*** (0.025)	0.027 (0.021)	0.126*** (0.025)
Availability of a porter service	0.867*** (0.052)	0.556*** (0.052)	0.879*** (0.052)	0.556*** (0.052)	0.873*** (0.053)	0.553*** (0.052)
Distance to the nearest school (km)	-0.010*** (0.002)	-0.008 (0.2)	-0.010*** (0.002)	-0.001 (0.002)	-0.010*** (0.002)	-0.001 (0.002)
Distance to the nearest hospital (km)	0.055*** (0.015)	0.061*** (0.015)	0.059*** (0.015)	0.060*** (0.015)	0.056*** (0.015)	0.058*** (0.015)
Distance to the nearest shops (km)	-0.187*** (0.032)	0.043 (0.033)	-0.187*** (0.032)	0.045 (0.032)	-0.187*** (0.032)	0.044 (0.033)
Log employment density	0.007 (0.013)	0.012 (0.013)	-0.001 (0.013)	0.011 (0.013)	0.002 (0.013)	0.008 (0.010)
Log shopping trips	0.074*** (0.010)	0.02*** (0.010)	0.078*** (0.010)	0.028*** (0.010)	0.076*** (0.010)	0.029*** (0.010)
R^2 *within*	0.1248	0.4197	0.1250	0.4197	0.1249	0.4213
R^2 *between*	0.4888	0.7525	0.4836	0.7523	0.4865	0.7528
R^2 *overall*	0.4544	0.7234	0.4526	0.7233	0.4529	0.7239

Legend: RCS: repeated cross-sectional data. PP: Pseudo panel data

*, **, *** indicate significance at 10%, 5% and 1% respectively. Values in parentheses are standard errors.

Table 8-3: Results of the first version of the HP models for the effect of the metro on rental listings of residential properties (REIDIN)

Catchment zone	0.5 km		1 km		1.5 km	
Data structure	**RCS**	**PP**	**RCS**	**PP**	**RCS**	**PP**
No. Observations	**81,248**	**823**	**81,248**	**823**	**81,248**	**823**
Model type	**First version of HP model**	**First version of HP model**	**First version of HP model**	**First version of HP model**	**First version of HP model**	**First version of HP model**
Observation years	2007 – 2011					
Covariates						
Constant	5.685*** (0.371)	3.830*** (0.423)	6.185*** (0.440)	3.936*** (0.428)	6.360*** (0.458)	4.111*** (0.434)
Year 2010	-0.145*** (0.012)	-0.111*** (0.019)	-0.143*** (0.012)	-0.127*** (0.021)	-0.142*** (0.012)	-0.152*** (0.023)
Year 2011	-0.191*** (0.019)	-0.170*** (0.024)	-0.191*** (0.019)	-0.186*** (0.025)	-0.190*** (0.019)	-0.211*** (0.027)
Accessibility to a number of metro stations	0.041*** (0.012)	-0.009 (0.028)	0.016** (0.007)	0.027 (0.018)	0.010*** (0.004)	0.030*** (0.011)
Area (sq.m)	0.002*** (0.00002)	0.005*** (0.0002)	0.002*** (0.00002)	0.005*** (0.0002)	0.002*** (0.00002)	0.005*** (0.0002)
Age	0.007 (0.009)	-0.008 (0.008)	0.010 (0.009)	-0.008 (0.008)	0.010 (0.009)	-0.008 (0.008)
Availability of a gym	-0.054 (0.039)	0.028 (0.035)	-0.049 (0.039)	0.028 (0.035)	-0.047 (0.039)	0.028 (0.035)
Availability of a swimming pool	0.005 (0.026)	0.029 (0.035)	0.010 (0.026)	0.029 (0.035)	0.009 (0.026)	0.029 (0.035)
Availability of a porter service	0.689*** (0.076)	0.174** (0.076)	0.762*** (0.082)	0.183** (0.077)	0.784*** (0.083)	0.197** (0.077)
Distance to the nearest school (km)	-0.001 (0.003)	-0.002 (0.003)	0.0002 (0.003)	-0.003 (0.003)	0.0007 (0.003)	-0.002 (0.003)
Distance to the nearest hospital (km)	-0.016 (0.016)	0.027* (0.014)	-0.005 (0.016)	0.027* (0.014)	-0.006 (0.015)	0.022 (0.014)
Distance to the nearest shops (km)	-0.216*** (0.043)	0.018 (0.040)	-0.212*** (0.044)	0.010 (0.040)	-0.205*** (0.044)	0.016 (0.040)
Log employment density	0.200*** (0.046)	0.356*** (0.054)	0.126** (0.057)	0.342*** (0.055)	0.103* (0.059)	0.321*** (0.055)
Log shopping trips	0.054** (0.026)	0.135*** (0.028)	0.029 (0.029)	0.130*** (0.028)	0.018 (0.030)	0.121*** (0.028)
R^2 *within*	0.1039	0.2256	0.1039	0.2326	0.1039	0.2376
R^2 *between*	0.7979	0.9270	0.7961	0.9262	0.7660	0.9269
R^2 *overall*	0.6183	0.8838	0.6145	0.8835	0.6087	0.8844

Legend: RCS: repeated cross-sectional data. PP: Pseudo panel data

*, **, *** indicate significance at 10%, 5% and 1% respectively. Values in parentheses are standard errors.

Table 8-4: Results of the first version of the HP models for the effect of the metro on sale transactions of retail properties (RERA)

Catchment zone	**0.5 km**		**1 km**		**1.5 km**	
Data structure	**RCS**	**PP**	**RCS**	**PP**	**RCS**	**PP**
No. Observations	**3,419**	**336**	**3,419**	**336**	**3,419**	**336**
Model type	**First version of HP model**	**First version of HP model**	**First version of HP model**	**First version of HP model**	**First version of HP model**	**First version of HP model**
Observation years	2007 – 2011					
Covariates						
Constant	14.979*** (0.504)	14.845*** (0.510)	14.834*** (0.519)	14.822*** (0.501)	14.858*** (0.494)	14.795*** (0.478)
Year 2008	0.404*** (0.097)	0.286*** (0.092)	0.393*** (0.096)	0.290*** (0.090)	0.362*** (0.096)	0.283*** (0.092)
Year 2009	0.307*** (0.096)	0.261*** (0.090)	0.316*** (0.095)	0.267*** (0.088)	0.279*** (0.096)	0.261*** (0.090)
Year 2010	0.053 (0.099)	0.164 (0.102)	-0.012 (0.098)	0.055 (0.103)	-0.038 (0.099)	-0.115 (0.122)
Year 2011	-0.125 (0.100)	-0.013 (0.105)	-0.199* (0.098)	-0.117 (0.105)	-0.218** (0.100)	-0.282** (0.124)
Accessibility to a number of metro stations	0.356*** (0.052)	0.234** (0.108)	0.336*** (0.029)	0.245*** (0.058)	0.138*** (0.014)	0.175*** (0.037)
Area (sq.m)	0.005*** (0.0004)	0.004*** (0.0004)	0.006*** (0.0004)	0.004*** (0.0004)	0.006*** (0.0004)	0.004*** (0.0003)
Distance to the nearest highway (km)	-0.08*** (0. 01)	-0.03 (0.02)	-0.06*** (0.01)	-0.04** (0.02)	-0.04*** (0.01)	-0.03 (0.02)
Log employment density	-0.108** (0.048)	-0.110** (0.047)	-0.116** (0.049)	-0.105** (0.046)	-0.151*** (0.047)	-0.130*** (0.043)
Log shopping trips	-0.224*** (0.043)	-0.192*** (0.043)	-0.203*** (0.044)	-0.182*** (0.042)	-0.197*** (0.042)	-0.168*** (0.040)
R^2 within	0.1369	0.1132	0.1592	0.1749	0.1500	0.1437
R^2 between	0.6287	0.6515	0.6304	0.6631	0.6295	0.6795
R^2 overall	0.6120	0.6018	0.6140	0.6188	0.5868	0.6260

Legend: RCS: repeated cross-sectional data. PP: Pseudo panel data

*, **, *** indicate significance at 10%, 5% and 1% respectively. Values in parentheses are standard errors.

Table 8-5: Results of the second version of the HP models for the effect of the metro on property values

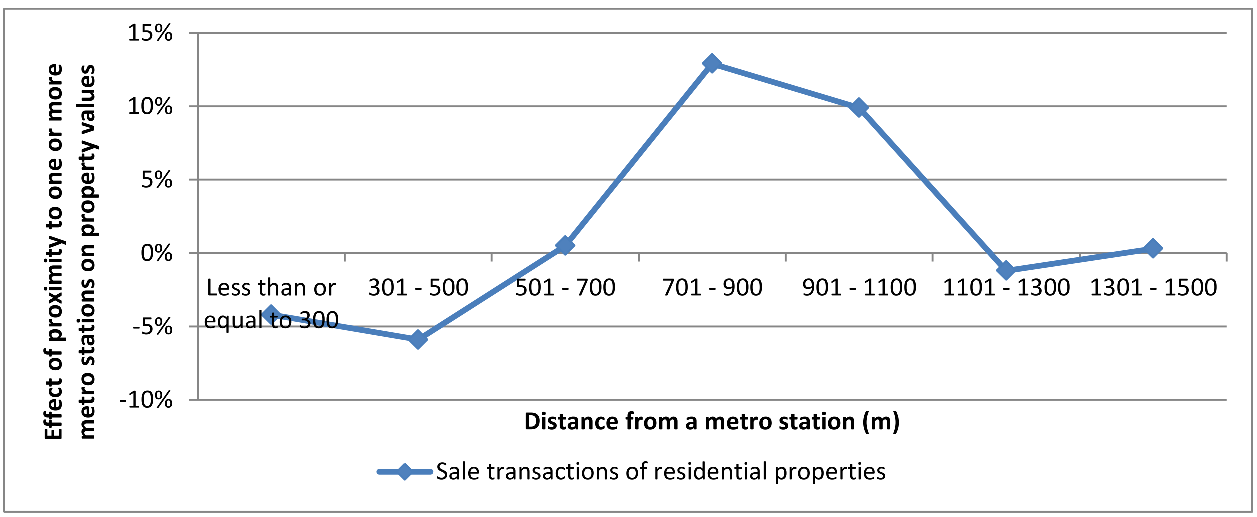
Dataset	Sale transactions of residential properties	Sale listings of residential properties	Rent listings of residential properties	Sale transactions of retail properties
Data structure	**RCS**	**RCS**	**RCS**	**RCS**
No. Observations	**39,308**	**165,978**	**81,248**	**3,419**
Observation years	2007-2011	2009-2011	2009-2011	2007-2011
Covariates				
Constant	11.992*** (0.119)	11.761*** (0.072)	5.413*** (0.468)	13.191*** (0.617)
Year 2008	0.309*** (0.009)	NA	NA	0.366*** (0.097)
Year 2009	0.265*** (0.010)	Reference year	Reference year	0.276*** (0.096)
Year 2010	0.219*** (0.014)	-0.128*** (0.008)	-0.152*** (0.012)	-0.045 (0.099)
Year 2011	0.182*** (0.015)	-0.244*** (0.014)	-0.201*** (0.019)	-0.234** (0.100)
Distance to a metro station				
Less or equal to 300 m	**-0.016 (0.025)**	**-0.004 (0.009)**	**0.086*** (0.022)**	**0.203*** (0.061)**
301 - 500 m	**-0.045** (0.020)**	**0.014** (0.007)**	**0.025* (0.015)**	**0.381*** (0.073)**
501 - 700 m	**0.015 (0.020)**	**0.044*** (0.006)**	**-0.013 (0.009)**	**0.395*** (0.071)**
701 - 900 m	**0.097*** (0.019)**	**0.018*** (0.006)**	**0.095*** (0.019)**	**0.745*** (0.075)**
901 - 1100 m	**0.075*** (0.018)**	**0.069*** (0.008)**	**0.034* (0.019)**	**0.378*** (0.091)**
1101 - 1300 m	**-0.016 (0.016)**	**0.054*** (0.011)**	**0.019 (0.022)**	**0.284*** (0.083)**
1301 - 1500 m	**-0.010 (0.019)**	**0.018 (0.022)**	**0.059 (0.081)**	**0.273** (0.122)**
Area (sq.m)	0.006*** (0.0001)	0.001*** (0.000001)	0.002*** (0.000002)	0.005*** (0.0003)
Age	NA	0.015** (0.007)	0.009 (0.009)	NA
Availability of a gym	NA	-0.098*** (0.028)	-0.066* (0.039)	NA
Availability of a swimming pool	NA	0.031 (0.021)	0.015 (0.027)	NA
Availability of a porter service	NA	0.873*** (0.052)	0.652*** (0.085)	NA

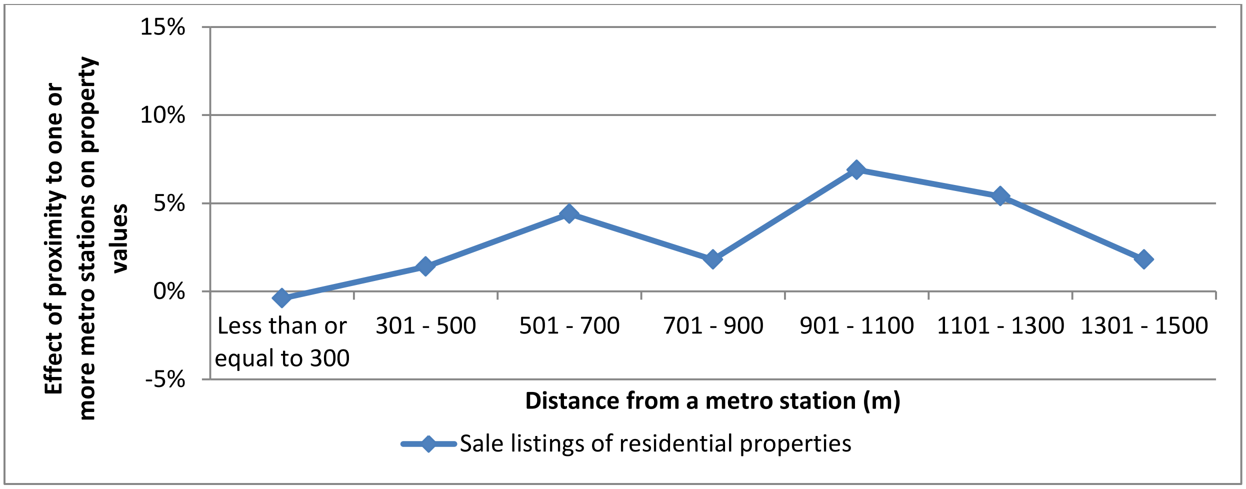
Distance to the nearest school (km)	-0.3*** (0.02)	-0.010*** (0.002)	-0.0007 (0.003)	NA
Distance to the nearest hospital (km)	0.1*** (0.02)	0.067*** (0.014)	-0.020 (0.016)	0.2*** (0.05)
Distance to the nearest shops (km)	0.1*** (0.02)	-0.203*** (0.032)	-0.227*** (0.044)	0.1 (0.01)
Distance to the nearest highway (km)	0.01*** (0.002)	NA	NA	0.02 (0.02)
Log employment density	0.149*** (0.011)	-0.008 (0.013)	0.230*** (0.060)	-0.097 (0.051)
Log shopping trips	0.021 (0.014)	0.081*** (0.010)	0.072** 0.030)	-0.186*** (0.040)
R^2 *within*	0.1933	0.1255	0.1045	0.1624
R^2 *between*	0.3570	0.4892	0.7958	0.7409
R^2 *overall*	0.3739	0.4623	0.6081	0.6950

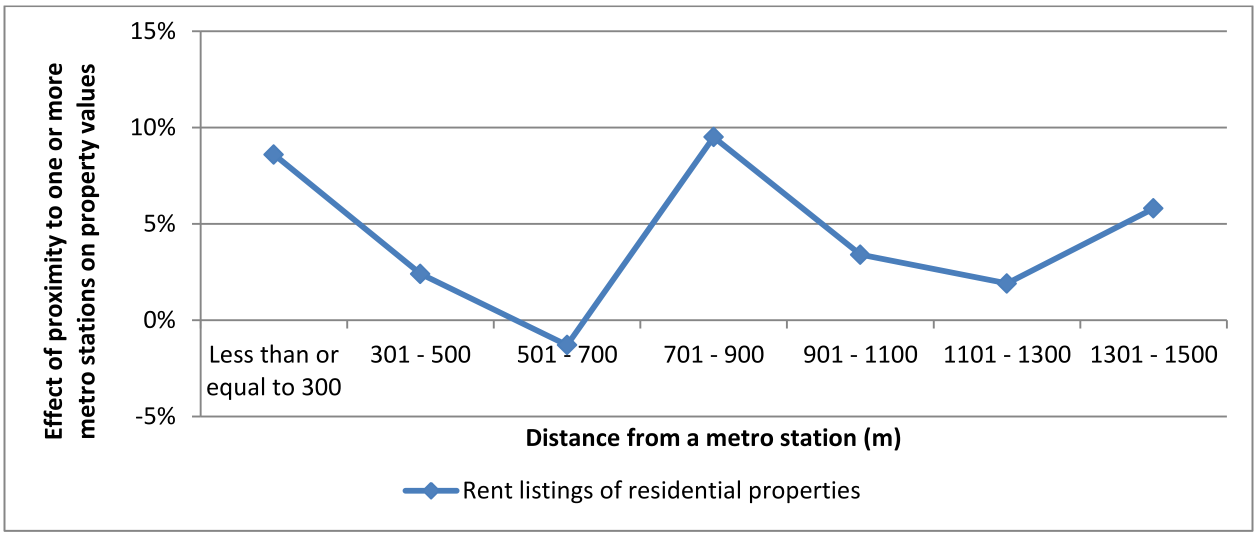
Legend : RCS: repeated cross-sectional data. PS: Pseudo panel data

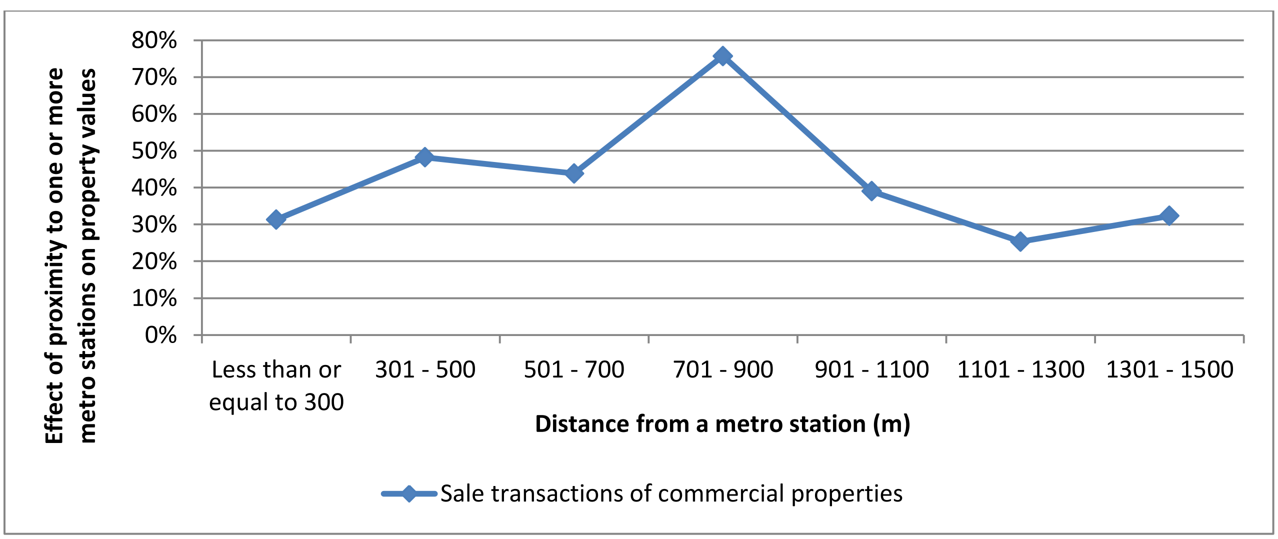
*, **, *** indicate significance at 10%, 5% and 1% respectively. Values in parentheses are standard errors. [a] annualized effect is calculated.

Figure 8-1: The effect of the Dubai Metro on the value of residential and retail properties at different distances, including significant and insignificant estimates




Effect of proximity to one or more metro stations on property values
15%
10%
5%
0%
-5%
Less than or equal to 300
301 - 500
501 - 700
701 - 900
901 - 1100
1101 - 1300
1301 - 1500
Distance from a metro station (m)
Sale listings of residential properties


Effect of proximity to one or more metro stations on property values
15%
10%
5%
0%
-5%
Less than or equal to 300
301 - 500
501 - 700
701 - 900
901 - 1100
1101 - 1300
1301 - 1500
Distance from a metro station (m)
Rent listings of residential properties


Effect of proximity to one or more metro stations on property values
80%
70%
60%
50%
40%
30%
20%
10%
0%
Less than or equal to 300
301 - 500
501 - 700
701 - 900
901 - 1100
1101 - 1300
1301 - 1500
Distance from a metro station (m)
Sale transactions of commercial properties

8.4 Conclusions

In this chapter, HP models are employed to estimate the effect of the Dubai Metro on property values. Property values were regressed on covariates that were expected to affect prices, including property and building attributes, distance to amenities, neighbourhood characteristics and distances to metro stations and highways. The effect of proximity to one or more metro stations was tested at the main catchment areas of 0.5 km, 1 km and 1.5 km using repeated cross-sectional data and pseudo panel data. In addition, for the repeated cross-sectional data, the effect of proximity to a metro station is examined at smaller catchment areas, for every 200 m starting at 0.3 km from a station.

Although the HP method provides an association between the regressors and the dependent variables, it has some drawbacks. In the presence of a causal relationship between the dependent variable and any of the covariates, the results from HP models are biased since they do not account for causal dependences. In addition, if significant variables are omitted from the model (intentionally or unintentionally), estimates generated by HP models are biased. This study has attempted to reduce such bias by adding as many covariates as possible that might be expected to affect property values, as well as by using repeated cross-sectional and pseudo panel data and controlling for time-invariant unobserved heterogeneity across properties. The validity of this approach is suggested by the fact that the findings from the HP models are in line with the findings from the DID models, which also control for causality and omitted variable bias.

Similar to the DID results, it is found that the impact of the metro on the sale transaction and listing values of residential properties is negative for properties close to a station (up to a maximum of 500 m), but is positive for properties located up to 1.1 km, reaching a positive peak for dwellings located at about 0.7 km. Nevertheless, the findings are mainly positive for the rental values of dwellings at a shorter radius, which indicates that the accessibility benefit to a metro station outweighs the impact of negative externalities on the value of rented properties in Dubai. Turning to the impact on the sale value of retail properties, a concave effect can be observed with only positive impacts due to proximity to a station across all distances up to a 1.5 km radius.

While the results simply estimate the effect of the metro alone, the analysis in the next chapter examines the transport system wide effect of operating the metro on the value of properties by testing for the impact on property values of the change in the generalized cost of travel.

Chapter 9. ESTIMATING THE EFFECT OF THE GENERALIZED COST OF TRAVEL ON PROPERTY VALUES

9.1 Introduction

As mentioned in the literature review section, there is a need in the literature to estimate the effect of transport on property values considering the change in travel times and costs resultant from transport system innovations. This study has attempted to address a part of this gap by measuring the change in the generalized cost of travel (GC) following the opening of the Dubai Metro, and its consequent effect on the value of properties.

To our knowledge, there is to date only one study which has attempted to estimate the effect on the value of properties due to changes in overall travel times and costs (Dewees, 1976). The differences between this study and ours are threefold. First, although Dewees (1976) attempted to find the effect of the overall travel costs on the value of properties, due to data limitations, the results are only available before the opening of the rail service. This study, however, tests for the effect of the change in the GC of travel by public and private transport with and without the metro. Second, while the previous study only examined the impact on properties located within one mile, this study estimates the impact on the value of properties located within different catchment areas of a metro. Third, this study uses records of property data for five years, covering the period before and after the opening of the metro, whereas Dewees (1976) uses data for one year before and one year after the opening of a railway.

As was explained in chapter 6, the GC value includes the monetary cost of travel and the value of time for public and private transport for each of the resident groups in a given community. It is therefore assumed that the perception of residents regarding the use of the metro, and the perceived benefit of accessibility to the metro, are already taken account of in the value of the GC of travel. This chapter is structured as follows. Section 9.2 explains the model design, the results for each of the datasets are presented and discussed in section 9.3, and a conclusion is offered in section 9.4.

9.2 Model specification

We test the effect of the generalized cost of travel before and after the opening of the metro Red Line using two model specifications in a hedonic pricing model structure. Using repeated cross-sectional data and pseudo panel data, the first two models estimate the effect of the change due to the opening of the metro in the GC of public transport as a whole (GC.PuT) (i.e. for the metro, buses and marine transport). Since the metro also affected the GC of private transport modes, however, the impact on property values of the weighted average generalized cost of travel (W.GC) for both private transport and public transport is also estimated using repeated cross-sectional data and pseudo panel data. The W.GC weighs the GC of travel of each mode by the trip share of that mode within the community where a property is located. For repeated cross-sectional data, the W.GC is calculated as shown below.

$$W.GC_{j(c)t} = \mathrm{SPuT}_{\mathrm{t,c}} GC.PuT_{j(c)t} + \mathrm{SPrT}_{\mathrm{t,c}} GC.PrT_{j(c)t}, \qquad (22)$$

Where:

$W.GC_{j(c)t}$ is the weighted generalized cost of travel to all communities in the emirate in time t for building j which holds a property i and is situated in community c

$\mathrm{SPuT}_{\mathrm{t,c}}$ is the share of public transport trips within community c in time t

$GC.PuT_{j(c)t}$ is the generalized cost of travel to all communities in the emirate by public transport services in time t for building j which holds a property i and is situated in community c

$\mathrm{SPrT}_{\mathrm{t,c}}$ is the share of private transport trips within community c in time t

$GC.PrT_{j(c)t}$ is the generalized cost of travel to all communities in the emirate by private transport in time t for building j which holds a property i and is situated in community c

Similarly, for pseudo panel data, the weighted average generalized cost is as follows:

$$W.GC_{j(c)t} = \mathrm{SPuT}_{\mathrm{t,c}} GC.PuT_{j(c)t} + \mathrm{SPrT}_{\mathrm{t,c}} GC.PrT_{j(c)t} \quad (23)$$

Where:

$W.GC_{j(c)t}$ is the weighted generalized cost of travel to all communities in the emirate in time t for pseudo panel group j situated in community c

$GC.PuT_{j(c)t}$ is the generalized cost of travel to all communities in the emirate by public transport services in time t for pseudo panel group j situated in community c

$\mathrm{GC.PrT}_{\mathrm{j(c)t}}$ is the generalized cost of travel by private transport to all communities in the emirate in time t for pseudo panel group j situated in community c

The opening of the Dubai Metro has led to changes in the GC of travel in almost all communities in Dubai. To distinguish between the effect on the value of properties located at different catchment areas to the metro (within 0.5 km, more than 0.5 km to 1 km, more than 1 km to 1.5 km and more than 1.5 km), the dummy variable D_j, which refers to the property location, is interacted with the generalized cost of travel from that location to all destinations in Dubai. The index "j" refers to the building in which a property is located for the models using repeated cross-sectional data, and the pseudo panel group for the models using pseudo panel data. Equations 24 to 27 present the considered models.

The effect of public transport GC using repeated cross-sectional data:

$$\ln y_{i(jc)t,} = \alpha + \alpha_t t + \sum_k \beta_{GC.PuT} GC.PuT_{j(c)t}.D_{jk} + \alpha_x X_{i(j)t} + \alpha_L L_{i(j)t} + \alpha_C C_{i(j)t} + \mu_j + \varepsilon_{(j)it,} \quad (24)$$

The effect of public transport GC using pseudo panel data:

$$\ln y_{j(c)t} = \alpha + \alpha_t t + \sum_k \beta_{GC.PuT} GC.PuT_{j(c)t}.D_{jk} + \alpha_x X_{jt} + \alpha_L L_{jt} + \alpha_C C_{jt} + \mu_j + \varepsilon_{jt} \quad (25)$$

The effect of weighted average GC using repeated cross-sectional data:

$$\ln y_{i(jc)t} = \alpha + \alpha_t t + \sum_k \beta_{W.GC.D} W.GC_{j(c)t}.D_{jk} + \alpha_x X_{i(j)t} + \alpha_L L_{i(j)t} + \alpha_C C_{i(j)t} + \mu_j + \varepsilon_{i(j)t} \quad (26)$$

The effect of weighted average GC pseudo panel data:

$$\ln y_{j(c)t} = \alpha + \alpha_t t + \sum_k \beta_{W.GC.D} W.GC_{j(c)t}.D_{jk} + \alpha_x X_{jt} + \alpha_L L_{jt} + \alpha_C C_{jt} + \mu_j + \varepsilon_{jt} \quad (27)$$

Where:

$y_{i(jc)t}$ is the value of property i located in building j and community c in time t

$y_{j(c)t}$ is the value of pseudo panel group j and community c in time t

D_j is a dummy variable that defines the catchment zone of a property to a metro station (within 0.5 km, more than 0.5-1 km, more than 1-1.5 km and beyond 1.5 km)

$X_{i(j)t}$ is a vector of property characteristics for property i located in building j in time t

X_{jt} is a vector of mean values of property characteristics for pseudo panel group j in time t

$L_{i(j)t}$ is a vector of location attributes for property i located in building j in time t (this includes distance to the nearest school, local shops, nearest highway)

L_{jt} is a vector of location attributes for pseudo panel group j in time t (same as above)

$C_{i(j)t}$ is a vector of the neighbourhood attributes for property i located in building j in time t (this is employment and shopping densities)

C_{jt} is a vector of neighbourhood attributes for pseudo panel group j in time t (same as above)

μ_j controls for time-invariant unobserved heterogeneity related to the building j (repeated cross-sectional data) and the pseudo panel group j (pseudo panel data)

These models have been designed to control for the possible endogeneity of the new metro. This can arise due to omitted variable bias since, for example, properties near the metro may happen to be in high-valued communities for reasons other than access to the metro, may be of higher quality compared to properties elsewhere and may be located in neighbourhoods with better air quality and less noise. Some of these unobserved factors are probably correlated with the motorized trip levels in the community in which a property is located and hence are correlated with the generalized cost of travel. If not accounted for, endogeneity can lead to inconsistent estimates of the regression parameters.

Similar to the other two models presented in chapters 7 and 8, this potential bias has been controlled for by using two data structures (repeated cross-sectional data and pseudo panel data) that are based on clustering and grouping observations to location and similar area categories. In addition, the factors that affect the choice of metro location (employment density and shopping commuting trips) are also included in the models in order to control for the fundamental differences in the communities adjacent to (within 1.5 km of a station) and distant from (beyond 1.5 km) metro stations. The results are discussed in the next section for each of the four property datasets.

9.3 Results

The results of the four models listed in this chapter are presented for each dataset (Table 9-1 to Table 9-4). The findings for the effect of the change in the GC.PuT due to the operations of the metro are compared to the results for the effect of the change in the W.GC. Similar to the HP models, the GC models reveal the average effect of the metro over 5 and 3 years for RERA and REIDIN datasets, respectively. We therefore also report the annualized effect of the change in the GC due to the operations of the metro in Appendix E.

9.3.1 Sale transactions of residential properties

Table 9-1 shows the results for sale transactions of residential properties, and reveals that the results using repeated cross-sectional data and pseudo panel data are not always comparable. The findings using the repeated cross-sectional data show that the reduction in the GC.PuT reduces property values located within 0.5 km of a metro station by 0.2%. This is probably due to the increased negative externalities for dwellings very close to a station. In comparison to the findings from previous studies and the earlier models, this result is not surprising. Nonetheless, no noticeable difference is found on the value of properties within the same catchment area using pseudo panel data.

While the results are not significant for dwellings located between 0.5 km and 1 km of a station, the value of residential properties increases with a decrease in the GC of travel beyond 1 km (using repeated cross-sectional data) and beyond 1.5 km (using pseudo panel data). The results also indicate that for every one Dirham (the currency in Dubai) reduction in the W.GC the value of properties located between 1 km and 1.5 km, and beyond 1.5 km is increased by about 0.3% and 2%, respectively (using repeated cross-sectional data). The findings from the model using pseudo panel data are larger in magnitude and reveal that the values of adjacent and distant properties are increased due to the reduction by one Dirham in the W.GC after the metro by a range of 3.2% to 5.3%. The difference in the estimates using the two data structures is probably due to measurement error in the pseudo panel data, which is explained further in the next chapter.

Turning to the effect of other variables on the value of dwellings and similar to the results in chapter 8, the coefficients indicate that proximity to schools increases the sale value of residential properties by 10% to 20%, whereas proximity to hospitals, shops and a highway reduces values. In addition, the models using repeated cross-sectional data suggest that the greater the employment density and the number of shopping trips, the higher the value of dwellings by a maximum of 9.5% and 6%, respectively. Comparing the fit in both data structures, the model using repeated cross-sectional data explains between 40% and 41% of the variation in property values, whereas the model using pseudo panel data explains between 37% and 38% of the variation.

9.3.2 Sale listings of residential properties

The findings for the effect of the generalized cost of travel on the sale listings of residential properties are presented in Table 9-2. For the model using repeated cross-sectional data, a one

Dirham reduction in the GC by public transport increases property values by 0.1% and 0.2% for properties located within 0.5 km and from 0.5 km to 1 km, respectively. There is no noticeable effect, however, on the value of dwellings within 1 km using pseudo panel data.

For properties located between 1 km and 1.5 km from a metro station, on the other hand, while the change in the GC has no effect when using repeated cross-sectional data, there is an increase in property values of 0.4% when using pseudo panel data. The latter result is not reasonable and is probably a biased estimated due to measurement error in the pseudo panel data. Furthermore, the value of dwellings located more than 1.5 km from a station increases by 0.3% with a unit decrease in the GC.PuT using both repeated cross-sectional data and pseudo panel data. The results are more consistent for the higher catchment area because in these areas the sample size also increases.

Moving to the findings for the effect of the W.GC, the estimates suggest an increase in the value of dwellings across Dubai due to a decrease in the W.GC in a range of 2.1% to 3.7%, using repeated cross-sectional data. On the other hand, the results are mixed using pseudo panel data. No impact on the value of properties located within 1 km is observed, but there are significant impacts for properties located further away. In contrast to the results using repeated cross-sectional data, there is an unexpected positive effect of an increase in the W.GC on the value of dwellings located within 1 km to 1.5 km. However, the coefficient is negative, as expected, for properties beyond 1.5 km (-2.7%).

The results also reveal substantial effects of other variables on sale listing values of residential properties. Not surprisingly, the asking value of dwellings increases by between 0.1% and 0.4% with a one square metre increase in property area, Regarding increased property age, there is no effect on values when using pseudo panel data, but the cross-sectional data suggests an increase in the value of dwellings with an increase in age. Similar to the findings in chapter 8, an explanation for this is that a number of properties listed for sale were not yet ready at the time of sale and therefore the values were higher after the properties were ready for occupation.

The coefficients for the building characteristics reveal that the availability of a gym reduces sale asking values by a minimum of 10%. With regards to the impact of the availability of a swimming pool, no effect is evident using repeated cross-sectional data, but the models using pseudo panel data show an increase in values of around 9%. In addition, a porter service in the building enhances values significantly in a range of between 44% and 78%.

The findings also indicate that proximity to schools increases property sale values, whereas proximity to hospitals reduces values. Additionally, the proximity of a property to shops increases values using repeated cross-sectional data, but no effect is found using pseudo panel data. Finally, the majority of the findings indicate no effect due to employment density but an increase in the number of shopping trips in a neighbourhood enhances sale listing values. The models using the repeated cross-sectional data explain between 39% and 47% of the variation in the data, whereas estimates using the pseudo panel dataset explain between 75% and 86% of the variation.

9.3.3 Rental listings of residential properties

Turning to the effect of the change in the GC on rental listing values, the results are comparable across models and catchment areas. The findings indicate that a unit reduction in the GC.PuT increases the value for adjacent and distant properties by between 0.4% and 0.6% using repeated cross-sectional data. The results using pseudo panel data, meanwhile, suggest that the reduction in GC.PuT increases rental values by 0.2% to 0.3% for residential properties located within 1.5 km but no effect is found for dwellings located further than 1.5 km from a station.

Similarly, the estimates suggest that the lower the weighted average generalized cost of travel, the higher the rental asking values for dwellings within 1.5 km, by between 0.7% and 0.9% and 0.5% to 0.6%, using repeated cross-sectional data and pseudo panel data, respectively. While the null hypothesis of no-effect cannot be rejected for properties located beyond 1.5 km from a station using pseudo panel data, the results using repeated cross-sectional data suggest an increase in the rental value of dwellings located more than 1.5 km from a station due to an increase in the W.GC.

Similar to the results of previous models, a one square metre increase in property area increases rental values by between 0.2 - 0.5%. For the majority of models, property age does not affect the rental values except for the model estimating the effect of the GC.PuT using repeated cross-sectional data, where there is a positive increase in rental value with increasing age.

Using repeated cross-sectional data, the findings indicate that the availability of a gym reduces the rental value of dwellings by at least 5.3% which may be due to the increase in the level of noise and vibration in the building, whereas no noticeable effect is found for the model using pseudo panel data. In addition, the findings indicate no effect of a swimming pool on rental values; however the availability of a porter service enhances values significantly by a maximum of 18% and 72% using repeated cross-sectional and pseudo panel data, respectively.

The results also reveal no effect of accessibility to schools and hospitals on the rental value of residential properties, except for the model using pseudo panel data, where proximity to a hospital reduces values by about 0.4%. This can be explained by the increase in the level of noise around a hospital which can disturb residents. While the findings indicate that accessibility to shops enhances values from 17% to 33% using repeated cross-sectional data, the null hypothesis of no-effect cannot be rejected using pseudo panel data. Similar to previous models, the results show that the higher the employment density and the number of shopping trips, the higher the rental values by a maximum of 44% and 15%, respectively. The explanatory power of the models ranges between 62% and 88%.

9.3.4 Sales transactions of retail properties

The results for retail properties are comparable between the two data structures for the model testing for the effect of the change in GC.PuT. The findings are not similar between the two data structures, however, for the model testing for the W.GC.

The results indicate that a reduction in the GC of travel by public transport increases the value of adjacent and distance retail properties by between 1.2% and 2.5% using repeated cross-sectional data. Similarly, the findings using pseudo panel data indicate an increase in value of between 1% and 2.4%.

While a reduction in the GC by public transport increases retail property values, the model examining the effect of change in the W.GC reveals the opposite. The estimates using repeated cross-sectional data suggest that a one Dirham increase in the W.GC increases values of adjacent and distant retail properties by about 11% and 8%, respectively. The results are insignificant when using pseudo panel data, however.

The findings from the two types of models indicate that although proximity to public transport services has a positive effect on the value of retail properties, property values are higher in areas with a higher weighted average GC. This can be explained through two main factors. First, due to the fact that the metro serves areas with high market attractiveness (e.g. higher employment and shopping densities, as explained in chapter 6), the treated retail properties are situated in locations with higher trip rates and congestion levels. This implies that the impact of the area characteristics on property values is more significant the effect of the change in the W.GC. Previous literature has also suggested that railways located in more congested areas have a greater effect on the value of properties (Cervero, 2003; Clower and Weinstein, 2002; Duncan, 2008). Second, this study (models in chapters 7 and 8) reveals a highly positive impact of the metro (at least 30%) on the value of retail properties located in close proximity to metro stations, which implies that the metro effect on property values supersedes the impact of the W.GC.

With regards to the effect from other variables, proximity to shops appears to have no significant impact, whereas a one kilometre reduction in the distance to a highway increases the value of retail properties by up to 7%. The findings also suggest that increases in employment density and shopping trips reduce the value of retail properties by a minimum of 12% and 8.3%, respectively. Analysing the goodness of fit measures using the two data structures, the models explain between 51% and 61% of the variation in the data using repeated cross-sectional data, and between 65% and 66% using pseudo panel data.

Table 9-1: Results of GC models for the effect of the metro on sale transactions of residential properties (RERA)

GC type	**GC of public transport**		**Weighted average GC**	
Data structure	**RCS**	**PP**	**RCS**	**PP**
No. Observations	**39,308**	**3,344**	**39,308**	**3,344**
Observation years	2007-2011			
Covariates				
Constant	12.008*** (0.116)	13.035*** (0.148)	12.093*** (0.117)	13.693*** (0.203)
Year 2008	0.314*** (0.009)	0.474*** (0.025)	0.316*** (0.009)	0.480*** (0.025)
Year 2009	0.277*** (0.010)	0.375*** (0.027)	0.287*** (0.010)	0.384*** (0.027)
Year 2010	0.267*** (0.010)	0.352*** (0.028)	0.285*** (0.011)	0.479*** (0.038)
Year 2011	0.232*** (0.010)	0.407*** (0.028)	0.252*** (0.011)	0.533*** (0.038)
GC (0.5 km)	**0.002*** (0.0005)**	**0.0001 (0.001)**	**-0.002 (0.003)**	**-0.038*** (0.009)**
GC (>0.5-1 km)	**-0.0002 (0.0005)**	**0.002 (0.001)**	**-0.0007 (0.002)**	**-0.032*** (0.009)**
GC (>1-1.5 km)	**-0.0003** (0.0002)**	**0.002 (0.001)**	**-0.003*** (0.0006)**	**-0.035*** (0.010)**
GC (>1.5 km)	**-0.004*** (0.0004)**	**-0.003*** (0.0009)**	**-0.020*** (0.003)**	**-0.053*** (0.010)**
Area (sq.m)	0.006*** (0.00006)	0.003*** (0.0001)	0.006*** (0.00006)	0.003*** (0.0001)
Distance to the nearest school (km)	-0.2*** (0.02)	-0.1*** (0.02)	-0.2*** (0.02)	-0.1*** (0.02)
Distance to the nearest hospital (km)	0.04** (0.02)	0.03 (0.02)	0.06*** (0.02)	0.08*** (0.02)
Distance to the nearest shops (km)	0.09*** (0.02)	0.06** (0.03)	0.1*** (0.02)	0.1*** (0.03)
Distance to the nearest highway (km)	0.02*** (0.001)	0.01** (0.003)	0.01*** (0.001)	0.001 (0.003)
Log employment density	0.095*** (0.012)	0.029** (0.014)	0.077*** (0.013)	0.096*** (0.012)
Log shopping trips	0.060*** (0.014)	0.007 (0.016)	0.063*** (0.015)	-0.033** (0.016)
R^2 *within*	0.1906	0.1117	0.1903	0.1066
R^2 *between*	0.3878	0.4239	0.3903	0.4356
R^2 *overall*	0.3968	0.3698	0.4141	0.3761

Legend: RCS: repeated cross-sectional data. PP: Pseudo panel data

*, **, *** indicate significance at 10%, 5% and 1% respectively. Values in parentheses are standard errors

Table 9-2: Results of GC models for the effect of the metro on sale listings of residential properties (REIDIN)

GC type	**GC of public transport**		**Weighted average GC**	
Data structure	**RCS**	**PP**	**RCS**	**PP**
No. Observations	**165,978**	**2,288**	**165,978**	**2,288**
Observation years	2009 - 2011			
Covariates				
Constant	12.054*** (0.070)	11.900*** (0.087)	12.672*** (0.083)	12.151*** (0.102)
Year 2010	-0.120*** (0.007)	-0.051*** (0.013)	-0.077*** (0.007)	-0.050*** (0.011)
Year 2011	-0.236*** (0.013)	-0.244*** (0.017)	-0.199*** (0.014)	-0.243*** (0.016)
GC (0.5 km)	**-0.001*** (0.0005)**	**-0.0002 (0.0007)**	**-0.024*** (0.002)**	**-0.0007 (0.002)**
GC (>0.5-1 km)	**-0.002*** (0.0004)**	**-0.000003 (0.0007)**	**-0.033*** (0.002)**	**-0.0010 (0.002)**
GC (>1-1.5 km)	**0.0007 (0.0006)**	**0.004*** (0.001)**	**-0.021*** (0.003)**	**0.006* (0.003)**
GC (>1.5 km)	**-0.003*** (0.0002)**	**-0.003*** (0.0005)**	**-0.037*** (0.002)**	**-0.027*** (0.003)**
Area (sq.m)	0.001*** (0.00001)	0.004*** (0.0001)	0.001*** (0.000001)	0.004*** (0.00009)
Age	0.013*** (0.007)	-0.002 (0.007)	0.019*** (0.007)	-0.0008 (0.007)
Availability of a gym	-0.126*** (0.026)	-0.099*** (0.034)	-0.159*** (0.027)	-0.140*** (0.035)
Availability of a swimming pool	0.015 (0.020)	0.091*** (0.020)	-0.015 (0.021)	0.086*** (0.024)
Availability of a porter service	0.783*** (0.048)	0.457*** (0.050)	0.579*** (0.051)	0.441*** (0.052)
Distance to the nearest school (km)	-0.010*** (0.002)	-0.002 (0.002)	-0.015*** (0.002)	-0.007*** (0.002)
Distance to the nearest hospital (km)	0.053*** (0.014)	0.051*** (0.015)	0.025* (0.015)	0.012 (0.016)
Distance to the nearest shops (km)	-0.216*** (0.030)	-0.019 (0.032)	-0.180*** (0.032)	-0.031 (0.033)
Log employment density	-0.0004 (0.012)	-0.014 (0.014)	0.082*** (0.014)	-0.013 (0.015)
Log shopping trips	0.069*** (0.009)	0.029*** (0.014)	0.010 (0.011)	0.027** (0.011)
R^2 *within*	0.1256	0.4132	0.1278	0.4261
R^2 *between*	0.5311	0.7892	0.4331	0.7819
R^2 *overall*	0.4720	0.7562	0.3948	0.7523

Legend: RCS: repeated cross-sectional data. PP: Pseudo panel data

*, **, *** indicate significance at 10%, 5% and 1% respectively. Values in parentheses are standard errors

Table 9-3: Results of GC models for the effect of the metro on rental listings of residential properties (REIDIN)

GC type	**GC of public transport**		**Weighted average GC**	
Data structure	**RCS**	**PP**	**RCS**	**PP**
No. Observations	**81,248**	**823**	**81,248**	**823**
Observation years	2009 - 2011			
Covariates				
Constant	6.057*** (0.399)	3.532*** (0.476)	3.647*** (0.434)	3.745*** (0.448)
Year 2010	-0.196*** (0.013)	-0.140*** (0.020)	-0.175*** (0.013)	-0.116*** (0.019)
Year 2011	-0.251*** (0.020)	-0.199*** (0.024)	-0.214*** (0.020)	-0.175** (0.023)
GC (0.5 km)	**-0.006*** (0.0008)**	**-0.003*** (0.0008)**	**-0.007*** (0.002)**	**-0.006*** (0.002)**
GC (>0.5-1 km)	**-0.004*** (0.0006)**	**-0.002*** (0.0008)**	**-0.009*** (0.002)**	**-0.005** (0.002)**
GC (>1-1.5 km)	**-0.006*** (0.001)**	**-0.003*** (0.001)**	**-0.008*** (0.002)**	**-0.006** (0.002)**
GC (>1.5 km)	**-0.006*** (0.0007)**	**-0.0008 (0.001)**	**0.014*** (0.002)**	**-0.003 (0.003)**
Area (sq.m)	0.002*** (0.00002)	0.005*** (0.0002)	0.002*** (0.00002)	0.005*** (0.0002)
Age	0.015* (0.009)	-0.008 (0.008)	-0.002 (0.009)	-0.007 (0.008)
Availability of a gym	-0.086** (0.039)	0.024 (0.035)	-0.053 (0.039)	0.024 (0.035)
Availability of a swimming pool	0.030 (0.027)	0.039* (0.024)	-0.017 (0.039)	0.039* (0.024)
Availability of a porter service	0.712*** (0.081)	0.179** (0.076)	0.578*** (0.077)	0.161** (0.076)
Distance to the nearest school (km)	-0.001 (0.003)	-0.002 (0.002)	0.002 (0.003)	-0.004 (0.002)
Distance to the nearest hospital (km)	0.023 (0.017)	0.043*** (0.015)	0.045** (0.018)	0.040*** (0.015)
Distance to the nearest shops (km)	-0.327*** (0.046)	0.007 (0.042)	-0.167*** (0.045)	0.005 (0.041)
Log employment density	0.170*** (0.050)	0.393*** (0.059)	0.438*** (0.054)	0.380*** (0.057)
Log shopping trips	0.079*** (0.029)	0.154*** (0.029)	0.148*** (0.028)	0.142*** (0.028)
R^2 *within*	0.1051	0.2340	0.1045	0.2260
R^2 *between*	0.7907	0.9285	0.8020	0.9290
R^2 *overall*	0.6205	0.8857	0.6303	0.8857

Legend: RCS: repeated cross-sectional data. PP: Pseudo panel data

*, **, *** indicate significance at 10%, 5% and 1% respectively. Values in parentheses are standard errors

Table 9-4: Results of GC models for the effect of the metro on sale transactions of retail properties (RERA)

GC type	**GC of public transport**		**Weighted average GC**	
Data structure	**RCS**	**PP**	**RCS**	**PP**
No. Observations	**3,419**	**336**	**3,419**	**336**
Observation years	2007-2011			
Covariates				
Constant	15.705*** (0.613)	15.364*** (0.698)	12.092*** (0.694)	14.230*** (0.715)
Year 2008	0.414*** (0.096)	0.269*** (0.091)	0.357*** (0.097)	0.270*** (0.091)
Year 2009	0.343*** (0.096)	0.266*** (0.090)	0.259*** (0.097)	0.243*** (0.089)
Year 2010	-0.052 (0.099)	0.050 (0.111)	0.016 (0.100)	0.210** (0.097)
Year 2011	-0.244** (0.100)	-0.129 (0.114)	-0.168* (0.101)	0.038 (0.099)
GC (0.5 km)	**-0.012*** (0.002)**	**-0.009** (0.004)**	**0.108*** (0.016)**	**0.017 (0.020)**
GC (>0.5-1 km)	**-0.017*** (0.002)**	**-0.010** (0.004)**	**0.113*** (0.018)**	**0.017 (0.022)**
GC (>1-1.5 km)	**-0.018*** (0.003)**	**-0.015*** (0.004)**	**0.103*** (0.019)**	**0.003 (0.024)**
GC (>1.5 km)	**-0.025*** (0.004)**	**-0.024*** (0.005)**	**0.078*** (0.022)**	**-0.029 (0.027)**
Area (sq.m)	0.005*** (0.0004)	0.004*** (0.0004)	0.006*** (0.0004)	0.004*** (0.0004)
Distance to the nearest highway (km)	-0.07*** (0.001)	-0.04* (0.03)	-0.02* (0.01)	-0.01 (0.09)
Log employment density	-0.127** (0.055)	-0.120** (0.054)	-0.156*** (0.059)	-0.120** (0.053)
Log shopping trips	-0.155** (0.052)	-0.100** (0.050)	-0.060 (0.055)	-0.083* (0.050)
R^2 *within*	0.1369	0.1263	0.1471	0.1341
R^2 *between*	0.6287	0.7361	0.6402	0.7233
R^2 *overall*	0.6120	0.6644	0.5078	0.6536

Legend: RCS: repeated cross-sectional data. PP: Pseudo panel data

*, **, *** indicate significance at 10%, 5% and 1% respectively. Values in parentheses are standard errors

9.4 Conclusions

In summary, the results for the effect of the change in GC due to the metro on the value of properties are mixed. Overall, the negative coefficients for the effect of the GC using public transport suggest that the value of residential and retail properties increases with a reduction in the travel times and costs associated with public transport following the opening of the metro (except for the sale value of residential properties located within 0.5 km of a station, which is due to negative externalities). In particular, it was found that with a unit decrease in the GC.PuT and W.GC the sale value of residential properties located across Dubai increased by between 0.1% and 0.3% for adjacent properties (i.e. within 1.5 km of a metro station), and 0.2% to 0.4% for distant properties, except for the actual sale values of dwellings located very close to a station.

With regards to the rental values, it was found that a unit reduction in the GC.PuT and W.GC increased rental values by between 0.4% and 0.6% across the emirate, except that an increase in the W.GC also increased the rental values for properties located more than 1.5 km from a station. The findings also reveal that a unit reduction in the GC.PuT enhances the value of retail properties across Dubai by between 1.2% and 1.8% for adjacent properties and 2.5% for distant properties. It was observed, however, that the value of retail properties increased with an increase in the W.GC.

Furthermore, the change in the GC.PuT due to the metro had a lesser effect on the value of properties located less than 1.5 km from a station compared to that for more distant properties. This indicates that although a reduction in the GC of public transport is appreciated across Dubai, distant properties are more sensitive to this change in the GC.PuT. In other words, changes in the GC of using public buses and marine services matters more for distant properties because there are less transport alternatives available for them. In contrast, properties close to a metro station (within 1.5 km) have access to more transport modes and services and although they still value accessibility to a metro station, the value of adjacent properties to metro stations are less responsive to changes in the GC.PuT.

Another explanation for this result can be related to the location choice of the metro. Since Dubai is a polycentric city with different employment destinations, and the Dubai Metro is located in areas with higher employment and shopping densities (**Error! Reference source not found.**), properties that are distant from metro stations are (in general) also more distant to major employment and shopping centres. Therefore, a change in the GC also results in a larger change in accessibility to major destinations for distant properties than for adjacent properties. This implies that a change in the GC is felt more for distant properties. Ryan (1999) argues that the elasticity of the change occurring due to that transport service is larger in polycentric cities and in distant areas.

Some differences were evident in the results generated using the two data structures, especially for adjacent properties to the metro. This is closely related to the possible bias in the models using pseudo panel data, as has occurred in the other models (chapters 7 and 8). The reasons for and implications of this are explained in more detail in the next chapter. In addition, the next chapter presents a summary of the key findings of this study and addresses the similarities and differences across property datasets, data structures and models.

Chapter 10. CONCLUSIONS AND WAY FORWARD

10.1 Introduction

Although research on the effect of railways on land and property values is mature, the results across case studies are very diverse and there is a lack of evidence for some cities. This study tests for the first time the effect of the recently opened Dubai Metro on the value of residential and retail properties. The findings of this research are not only of interest to Dubai, but also to cities with similar conditions, transport policies and residential travel behaviour and preferences. In particular, cities of the Gulf Co-operation Council (GCC) have similar characteristics to Dubai and many are building their first railway. Results from this study may offer guidance for the effect of the upcoming railways on property values and can provide a justification for partially funding railways using a part of the property value increase.

This chapter summarizes the study results in light of the research questions. These are to summarize the existing literature critically and to estimate the effect of the Dubai Metro on property values. The first objective was achieved by conducting an extensive review of the literature on the effect of railways on land and property values and supplementing this with a meta-analysis to explain the large range of variation in results across studies. Based on this, the main factors that affect property values were identified and as many of these as possible were obtained in the data used for this study.

The second objective of this study was to enrich, as much as possible, existing datasets on property values in Dubai. Four comprehensive datasets for residential and retail properties were therefore established, containing property sale or rental records, property and building attributes, location of property, neighbourhood attributes, transport data and data on proximity to amenities.

The third objective was achieved by using two data structures, different empirical methods and different measures of the accessibility offered by the metro. Starting with the datasets, four out of nine obtained datasets were suitable for analysis. These were the sale transactions of dwellings, sale listings of dwellings, rental listings of dwellings and sale transactions of retail properties. While the records on transactions were available for 5 years, records on listings were available for 3 years. The datasets consisted of repeated cross-sectional property records before and after the opening of the metro Red Line. Repeated cross-sectional data and pseudo panel data corrected for sources of bias were used in the study.

To examine the effect of the metro, we adopted three empirical methods. It is worth mentioning that the availability of the data and its structure can affect the choice of the empirical method. For example, the difference-in-differences (DID) method is valid only if data is available before and after a treatment and for treated and control groups. The study adopted the DID and HP methods since these represented the best fit to the available data and allowed sources of bias to be controlled for through the model structure as well as the data structure.

The effect of the Dubai Metro on land and property values was tested using three measures representing the accessibility offered by the metro. The first was a binary variable that distinguished the effect of proximity to a metro station on the value of nearby properties. The second examined the impact of proximity to one or more metro stations. In addition, there is a need in the literature to test for the effect of a transport system on the travel times and costs. Therefore, the third measure in this study is the effect on property values of the change in the generalized cost of travel due to the opening of the metro.

This final chapter is structured as follows. Section 10.2 highlights the differences in the data structure and the empirical methods and argues for the preferred ones. Section 10.3 summarizes the key results and compares the findings with the existing empirical work. Section 10.4 discusses the policy implications for decision makers and practitioners in the field. The contributions of this research are outlined in section 10.5, while section 10.6 discusses the limitations of the study and section 10.7 provides suggestions for future research.

10.2 Differences in the substance of the models and the data structure

The three empirical models (presented in chapters 7, 8 and 9) differ not only in their structure but also in the information each conveys. As discussed, the estimates from the full DID models are more consistent compared to those from the HP models, since the former captures causation rather than just the relationship between property values and the regressors. Nevertheless, we found comparable and consistent results across the empirical methods for the same dataset, which tends to indicate the consistency of the models used. While the first two models use similar variables to measure the impact as are used in the majority of previous empirical work, and examine the effect of the Dubai Metro alone, the models in chapter 9 estimate the impact of the change in the overall transport network due to the metro. Hence, one would expect the results from chapters 7 and 8 to be aligned and the results from chapter 9 to complement them.

The data structure used to estimate the impact of a transport system on the value of properties can also affect the generated results. As explained in chapter 5, the main advantage of creating pseudo panel data is to correct for unobserved heterogeneity. There is a possibility, however, of information loss that can lead to bias unless a sufficient number of records are grouped in one cohort. While the majority of previous studies in this field use repeated cross-sectional data, only a few have discussed its advantages. For example, Agostini and Palmucci (2008) and Koster et al. (2010) argue that in the absence of panel data, results are more consistent using repeated cross-sectional data compared to other data structures. In cases like the one under consideration here, since unobserved heterogeneity is controlled for in the models using repeated cross-sectional data, it can be contended that the findings from these models are more reliable.

The reasons for this are: first, that since the sample size in the repeated cross-sectional data is much higher than in the pseudo panel data (the latter contains no more than 10% of the original records), the results are likely to be more consistent. As explained in chapters 5 and 6, the arrangement of the data only allowed a relatively small number of observations to be retained in the pseudo panel data while also retaining a larger between-group than within-group variation.

Second, although the pseudo panel groups allows unobserved heterogeneity across cohorts to be controlled for, the unobserved factors for a given cohort can be time-variant, since each observation may still have some unique unobserved effect, and the loss of individual information can lead to measurement error. In this case, observations were grouped according to location and property size, and, in so doing, the unobserved factors related to the building (such as the number of floors, the quality of the finishes in the building, etc.) are fixed over time. In addition, the unobserved factors that may affect property values, such as the perception of the property buyer regarding the property size, are also assumed to be constant for the study time period. Nonetheless, the unobserved factors of a property that may affect its value (such as the amount of natural lighting) will still not be captured.

In the repeated cross-sectional data, there are a much larger number of records per cluster of properties compared to only one averaged value per cohort in the pseudo panel data. Since the sample size per cluster is larger, similar properties within a cluster may repeat over time, hence the unobserved factors can be treated as time-invariant and the unobserved factors related to a property are better controlled for. In fact, the results from the models using pseudo panel data varied compared to those using repeated cross-sectional data, with the majority of the findings from the pseudo panel data models suggesting less significant estimates. This was probably due to the reasons explained above.

It was also observed that the results using the two data structures were more comparable for the sale value of retail properties compared to the residential dwellings. This can be explained due to the following two reasons. First, the average values of treated and control properties using the two data structures (chapter 6) are most similar for the retail dataset. Second, the coefficients of variation for the retail property values using the two data structures are more consistent than those for the dwellings (Table 6-3). These suggest that the measurement error in the retail dataset is smaller, leading to more comparable estimates.

10.3 Summary of main findings

The previous section discussed the differences in the empirical methods and the impact of the data structure on the estimates. This section compares the key findings obtained using the preferred data structure (repeated cross-sectional data) and the preferred models (equations 17, 19, 21, 24 and 26) across datasets and estimation methods as well as with the results from other case studies. The main results are summarized in Table 10-1, and we also report the effect on an annualized basis and compare with other equivalent case studies.

As explained in the meta-analysis study (chapter 3) the effect of proximity to a rail station is lower in car-oriented cities and, given that the public transport share in Dubai is low, the findings for the effect of the Dubai Metro on the value of properties are likely to be comparable with the findings from case studies from other cities with low public transport share. In fact, similar results to those found in this study have been are evident in studies of railways in the USA (e.g. Agostini & Palmucci, 2008; Billings, 2011; Bowes & Ihlanfeldt, 2001; Clower & Weinstein, 2002; Duncan, 2008). Furthermore, almost all the findings in this study show concave effects of the Dubai Metro on the value of properties, whereby positive effects are highest at some distance away from a metro station, and similar results are also found in other case studies (e.g. Billings, 2011; Seo, Golub, & Kuby, 2014).

Overall, the findings indicate that the highest positive impact of the metro is on the value of retail properties (as is the case also for previous studies, which is mentioned in chapters 2 and 3), followed by the impact on the sale and rental listing values of dwellings and then the sale transaction value of dwellings. While the impact is always positive for retail properties and rented residential properties, the effect of proximity to a metro station (equations 17, 19 and 21) is positive on the sale value of dwellings located at walking distances to a station, but not too close to it. It is also observed that the radius of impact is wider for retail properties, which is contrary to the findings in the literature. This implies that the Dubai Metro is more valuable to commercial areas than it is for residential areas.

Comparing the results across methods, it was found that the estimates are comparable between the first two methods (equations 17, 19 and 21), whereas the radius of impact of the metro is wider for the models employing a HP method. With regards to the overall network effect due to the metro (equations 24 and 26), the results reveal positive effects in terms of improving the generalized cost of travel and enhancing property values, except for the sale value of residential properties located less than 0.5 km from a station, which is similar to the findings from the previous models.

Starting with equation 17 for dwellings located within 0.5 km of a station, it was found that while the sale values were reduced due to the metro (by an average -0.9% to -3.6% annually), the rental values increased (3.4% on annual basis). On the other hand, the results were positive for dwellings located within 1 km of a metro station; with an annual increase of 3.1%, 1.7% and 0.7% for the sale transaction, sale listing and rental listing values, respectively. In addition, the models reveal no significant effect of the metro on the value of residential properties situated within 1.5 km of a station, except for the sale asking value of dwellings (1.7% annually).

The negative effect of the metro is less for the sale listing value versus the transaction value of dwellings, and the impact radius of the metro on the earlier properties is also wider. These results imply a potentially larger positive effect of the metro on the sale transaction value of residential properties in the future. This is likely to occur after more metro lines are opened, the system stabilizes and the accessibility benefit to a metro station becomes more valuable to the residents, as was the case for the Metropolitan Atlanta Rapid Transit (Bowes and Ihlanfeldt, 2001) as well as the Bay Area Rapid Transit system in San Francisco (Clower and Weinstein, 2002; Landis et al., 1994).

The results produced here are comparable with those from previous studies using DID models on properties located within similar catchment areas as well as in cities with low public transport share. It is worth mentioning that no previous studies were found that used a DID method on properties located within 0.5 km to compare with our results, however results for a slightly larger catchment area are available. For example, in a study of the effect of light rail in Charlotte North Carolina USA, Billings (2011) finds no significant effect for single family homes located within half a mile (0.8 km) but a positive effect is estimated (1.7% on annual basis) on the value of condominiums situated within this catchment area, and estimates are also positive (a minimum of 0.5% on annual basis) for the value of residential properties located within one mile (1.6 km). The study by Agostini and Palmucci (2008) for the Santiago Metro also suggests that dwellings located within 1 km of a metro station increase in value by an annual average value of 1.1% during the construction stage.

The DID also suggests a positive effect of the metro on the value of retail properties located up to 1.5 km away, by a maximum of 16.8% on an annual basis. To the best of our knowledge, the only other study that uses a DID model to estimate the effect of rail on the value of commercial properties is that of Billings (2011), which finds no significant impact.

Turning to equations 19 and 21, the results reveal no noticeable effect of the metro on properties situated within 0.3 km, however the metro does appear to reduce the sale transaction value of properties situated between 0.3 km and 0.5 km. A few studies using HP methods also indicate a negative impact of railways on the value of dwellings located very close to a station. For example, Bowes and Ihlanfeldt (2001) find that residential properties located within a quarter of a mile (about 0.4 km) of a station experience a reduction in value of about -1.9% on annual basis, whereas properties located further away are positively affected. Although in an area with almost double the public transport share compared to Dubai, Du and Mulley (2006) find that, in some communities, values reduce within 0.2 km of a station but increase for properties located between 0.5 km and 1 km. This study also finds that the sale and rental listing values of dwellings located within this catchment area increase annually by 0.5% and 0.8%, respectively.

Moving to higher catchment areas, the results show that the peak positive impact of proximity to the Dubai Metro on the sale value of dwellings is at 0.7 km to 0.9 km (1.9% annually), whereas the peak is larger in magnitude and in radius for the sale listing values of dwellings (2.3% annually at 0.9 km to 1.1 km). It was also found that the impact on an annual basis of the metro on the rental listing value of dwellings peaks at two closer catchment areas: within 0.3 km (2.9%) and from 0.7 km to 0.9 km (3.2%). This is in line with the results from existing empirical work: Agostini and Palmucci (2008) (Santiago) estimate an annual increase in value from 0.9% to 1.9%, while Duncan (2008) (San Diego) finds a positive effect on an annual basis of 1.3% to 3.7% and Chen et al. (1997) (Portland) estimate an annual effect of 2.6%. Seo et al. (2014) (Phoenix) indicate slightly higher estimates for the annual effect of railways, which ranges between 5% and 6.1%.

Similar to the first models, the results for retail properties are larger; the values increase on an annual basis from about 4.1% to 14.9%, peaking at a distance of 0.7 km to 0.9 km of a station. Comparing with the two other case studies in cities with a low public transport share city, our estimates are within the range; Cervero (2003) (San Diego) finds that the proximity to a light railway station (within 0.8 km) increases the value of commercial properties on an annual basis from 24% to 30.3%, whereas Weinberger (2001) estimates that values increase between 0.8% and 1.3% annually.

Turning to the overall network effect of the metro, this study finds that a reduction in the GC by public transport increases the value of residential and retail properties across Dubai, except for the sale value of dwellings located within 0.5 km of a station. The effect of the W.GC is not similar across property types, however. While it is found that a reduction in the W.GC increases the value of residential properties, the opposite is true for retail properties. This result is sensible for retail properties since an increase in the travel times and costs can imply more travel activities within an area and hence a greater potential for retail activities.

Since this measure of the accessibility offered by the metro is new to the literature, it has not been possible to compare the results here directly with previous work. It was possible, however, to draw a comparison with the work of Dewees (1976), who attempted to examine the overall effect of a railway on the value of properties for a year before and a year after the opening of the system. Dewees, however, was only able to find the effect of travel times due to lack of data on costs after the opening. In addition, there were some inconsistent estimates across Dewees' models, hence only the findings from the model that is best related to this study are discussed here. Dewees' results indicate that the effect of reducing the travel time after the opening of the railway system is larger for longer distance trip, which is similar to the finding in this study, where there were slightly higher impacts of a change in the GC of travel on the value of properties located beyond 1.5 km from a station compared to closer properties.

Comparing these results with the expectations of real estate agents in Dubai (chapter 2) regarding the effect of the Dubai Metro on property values, the majority of the anticipations matched the estimated effect presented in this study. The agents anticipated either no or a negative effect of proximity to a metro station on the value of residential properties located very close to a station, which is similar to the results of this study. In addition, considering the average estimates over the years, the actual findings reveal a smaller positive impact of the metro on the value of dwellings located within 1 km of a station compared to the anticipated effect of 10% to 30%. With regards to the commercial properties, the real estate agents expected that the metro may increase values by up to 40%, which is consistent with the estimates using a DID method, but is far less than the result from the HP models. In summary, the majority of the results from this study reveal a lesser effect of the metro on the value of properties compared to the anticipated effect of the real estate agents. This may be due to the short study time period (i.e. post the announcement of the metro until two years of operations), which resulted in under-estimating the metro effect, in comparison to the longer anticipated time-frame effect of the Dubai Metro by the real estate agents.

Finally, comparing the goodness of fit of our models with the goodness of fit measures from the previous empirical work, our results are in the lower range. The models explain between 34% to about 70% of the variation in property data, whereas the range in the literate varies from 32% (e.g. FTA, 2000; McDonald and Osuji, 1995) to about 90% (e.g. Billings, 2011; Dubé et al., 2014; Gibbons and Machin, 2005), and the majority report values between 60% to 80% (e.g. Agostini and Palmucci, 2008; Ahlfeldt, 2013; Baum-Snow and Kahn, 2000; Dewees, 1976; Kim and Zhang, 2005; Koster et al., 2010; McMillen and McDonald, 2004; Zhang and Wang, 2013).

In addition, we find that the values for the goodness of fit are higher for the models using the commercial property dataset compared to the models using datasets on dwellings. This is related to the relatively smaller number of available physical attributes of residential properties that likely affect property values (such as number of bathrooms and the floor number where a property is located). Nevertheless, we have tried to overcome this by grouping records to similar clusters and controlling for unobserved heterogeneity within a cluster.

Table 10-1: A summary of the preferred model estimates (source: self-produced table)

Model type	**Full DID model**	**First version of the HP models**	**Full DID model**	**First version of the HP models**	**Full DID model**	**First version of the HP models**
Equation number	**17**	**19**	**17**	**19**	**17**	**19**
Catchment zone	**0.5 km**		**1 km**		**1.5 km**	
Data structure	**RCS**		**RCS**		**RCS**	
Dataset	Sale transactions of residential properties					
Metro effect	-0.090*** (0.014)	-0.086*** (0.013)	0.078*** (0.008)	0.019* (0.010)	0.018 (0.013)	0.012** (0.005)
Result on an annualized basis	-0.036	-0.017	0.031	0.004	0.007	0.002
R^2 *overall*	0.3361	0.3766	0.3467	0.3805	0.3855	0.3818
Dataset	Sale listings of residential properties					
Metro effect	-0.014** (0.006)	-0.014** (0.006)	0.026*** (0.005)	0.015*** (0.004)	0.026*** (0.004)	0.004*** (0.002)
Result on an annualized basis	-0.009	-0.005	0.017	0.005	0.017	0.001
R^2 *overall*	0.2739	0.4544	0.2588	0.4526	0.2228	0.4529

Dataset	Rent listings of residential properties					
Metro effect	0.051*** (0.013)	0.041*** (0.012)	0.010* (0.006)	0.016** (0.007)	0.008 (0.006)	0.010*** (0.004)
Result on an annualized basis	0.034	0.014	0.007	0.005	0.005	0.003
R^2 *overall*	0.4511	0.6183	0.4683	0.6145	0.4818	0.6087
Dataset	Sale transactions of retail properties					
Metro effect	0.394*** (0.054)	0.356*** (0.052)	0.420*** (0.037)	0.336*** (0.029)	0.405*** (0.036)	0.138*** (0.014)
Result on an annualized basis	0.158	0.071	0.168	0.067	0.162	0.028
R^2 *overall*	0.5080	0.6120	0.6065	0.6140	0.5366	0.5868

Model type	**Second version of the HP model**			
Equation number	**21**	**21**	**21**	**21**
Dataset	Sale transactions of residential properties	Sale listings of residential properties	Rental listings of residential properties	Sale transactions of retail properties
Data structure	**RCS**	**RCS**	**RCS**	**RCS**
Distance to a metro station				
Less or equal to 300 m	-0.016 (0.025)	-0.004 (0.009)	0.086*** (0.022)	0.203*** (0.061)
301 - 500 m	-0.045** (0.020)	0.014** (0.007)	0.025* (0.015)	0.381*** (0.073)
501 - 700 m	0.015 (0.020)	0.044*** (0.006)	-0.013 (0.009)	0.395*** (0.071)
701 - 900 m	0.097*** (0.019)	0.018*** (0.006)	0.095*** (0.019)	0.745*** (0.075)
901 - 1100 m	0.075*** (0.018)	0.069*** (0.008)	0.034* (0.019)	0.378*** (0.091)
1101 - 1300 m	-0.016 (0.016)	0.054*** (0.011)	0.019 (0.022)	0.284*** (0.083)
1301 - 1500 m	-0.010 (0.019)	0.018 (0.022)	0.059 (0.081)	0.273** (0.122)
Result on an annualized basis[a]				
Less or equal to 300 m	**-0.003**	**-0.001**	**0.029**	**0.041**
301 - 500 m	**-0.009**	**0.005**	**0.008**	**0.076**
501 - 700 m	**0.003**	**0.015**	**-0.004**	**0.079**
701 - 900 m	**0.019**	**0.006**	**0.032**	**0.149**
901 - 1100 m	**0.015**	**0.023**	**0.011**	**0.076**
1101 - 1300 m	**-0.003**	**0.018**	**0.006**	**0.057**
1301 - 1500 m	**-0.002**	**0.006**	**0.020**	**0.055**
R^2 *overall*	0.3739	0.4623	0.6081	0.6950

Equation number	24	26
GC type	GC of public transport	Weighted average GC
Data structure	RCS	RCS
Dataset	Sale transactions of residential properties	
	Metro effect	
GC (0.5 km)	0.002*** (0.0005)	-0.002 (0.003)
GC (>0.5-1 km)	-0.0002 (0.0005)	-0.0007 (0.002)
GC (>1-1.5 km)	-0.0003** (0.0002)	-0.003*** (0.0006)
GC (>1.5 km)	-0.004*** (0.0004)	-0.020*** (0.003)
	Result on an annualized basis	
GC (0.5 km)	0.0004	-0.0004
GC (>0.5-1 km)	-0.00004	-0.00014
GC (>1-1.5 km)	-0.00006	-0.0006
GC (>1.5 km)	-0.0008	-0.004
R^2 *overall*	0.3968	0.4141
Dataset	Sale listings of residential properties	
	Metro effect	
GC (0.5 km)	-0.001*** (0.0005)	-0.024*** (0.002)
GC (>0.5-1 km)	-0.002*** (0.0004)	-0.033*** (0.002)
GC (>1-1.5 km)	0.0007 (0.0006)	-0.021*** (0.003)
GC (>1.5 km)	-0.003*** (0.0002)	-0.037*** (0.002)
	Result on an annualized basis	
GC (0.5 km)	-0.0003	-0.008
GC (>0.5-1 km)	-0.0007	-0.011
GC (>1-1.5 km)	0.0002	-0.007
GC (>1.5 km)	-0.001	-0.012
R^2 *overall*	0.4720	0.3948
Dataset	Rental listings of residential properties	
	Metro effect	
GC (0.5 km)	-0.006*** (0.0008)	-0.007*** (0.002)
GC (>0.5-1 km)	-0.004*** (0.0006)	-0.009*** (0.002)
GC (>1-1.5 km)	-0.006*** (0.001)	-0.008*** (0.002)
GC (>1.5 km)	-0.006*** (0.0007)	0.014*** (0.002)
	Result on an annualized basis	
GC (0.5 km)	-0.002	-0.002
GC (>0.5-1 km)	-0.001	-0.003
GC (>1-1.5 km)	-0.002	-0.003
GC (>1.5 km)	-0.002	0.005
R^2 *overall*	0.6205	0.6303

Dataset	Sale transactions of retail properties	
	Metro effect	
GC (0.5 km)	-0.012*** (0.002)	0.108*** (0.016)
GC (>0.5-1 km)	-0.017*** (0.002)	0.113*** (0.018)
GC (>1-1.5 km)	-0.018*** (0.003)	0.103*** (0.019)
GC (>1.5 km)	-0.025*** (0.004)	0.078*** (0.022)
	Result on an annualized basis	
GC (0.5 km)	-0.002	0.022
GC (>0.5-1 km)	-0.003	0.023
GC (>1-1.5 km)	-0.004	0.021
GC (>1.5 km)	-0.005	0.016
R^2 *overall*	0.6120	0.5078

Legend : RCS: repeated cross-sectional data. PS: Pseudo panel data

*, **, *** indicate significance at 10%, 5% and 1% respectively. Values in parentheses are standard errors. [a] annualized effect is calculated.

10.4 Policy implications

The results of this study have some policy implications that are discussed here. As was explained in chapter 2, a number of cities have implemented a value capture mechanism to recover at least a part of the monetary benefit on the value of properties as a source to fund transport systems. Given the results of this study, a value capture mechanism could be explored for rented dwellings and retail properties, or for all properties that experience a positive effect from the Dubai Metro. A compensation mechanism could also be considered to alleviate the negative effect on property values, where present.

There are two main factors supporting the desire to claim at least a part of the property value increase due to the metro. First, the positive effects, albeit at different distances and in different orders of magnitudes, are reasonable and within the range of estimates generated in previous studies. Second, since the public did not have a choice on the location of the Dubai Metro, owners of properties located within the catchment area of a metro station benefit from the metro without contributing towards any of its costs. Charging at least a part of the value increase towards the costs of transport systems could be fair to owners of properties that either experienced a negative effect or are located in other parts of Dubai and were not offered any accessibility benefit.

It is worth mentioning that the results of this study reveal the average effect of the elevated part of the Dubai Metro on the value of properties. In theory, different results would be expected for the effect of proximity to metro stations that are underground, since the environmental conditions around such stations (e.g. the levels of noise and pollution due to the metro) are different, although properties near underground stations may suffer more from vibration. The policy implications suggested in this study should therefore be considered with caution in respect to the underground part of the Dubai Metro.

Nevertheless, two value capture policies can be explored. The first is to charge owners of properties that experience an increase in value a levy of at least a part of the average value increase, according to the rates obtained in this study. It is worth mentioning that the estimates are the average effect of the metro on property values rather than location-specific effects. The second is to employ a spatially varying model to obtain the impact at each location. Although the second option is fairer, it requires a more scattered property dataset (especially post-operations), which unfortunately was not available when this study was conducted. With regards to the value capture strategy, it is suggested that the partnership deals or the endowment funding scheme (introduced in chapter 2) are more suitable, justified and reasonable for Dubai, compared to charges and taxes, for the following reasons. First, they are a one-off time payment and, in a tax-free city, are more likely to be accepted by the public. Second, the payments are directly related to the benefit realized due to the transport system.

The above discussion also implies the need for policies related to the funding of future metro systems in Dubai. In a tax-free city, however, the public is not familiar with an expectation to contribute towards the services provided by the government,. Introducing a value capture scheme as a contribution to public transport systems has to be well justified to be accepted by the public, therefore. In addition, Dubai may consider other elements related to the value capture or compensation policy. These are the status of Dubai's economy, the competiveness of the emirate compared to neighbouring cities and the wider world, as well as the strategic growth plans for Dubai (Salon and Shewmake, 2012).

Additionally, the legal framework of property acquisition has to be modified to ensure that value capture is as easy in practice as the theory indicates (Enoch et al., 2005; Martínez and Viegas, 2012; Medda, 2012). For example, a decision is needed on the "payee" target (e.g. the owner or the user of the property) and the time of contribution (e.g. during construction of the metro, after the metro operations, during property sale).

The results of this study may also imply other considerations for Dubai. Since the majority of the models suggest positive effects of the metro on the value of properties, this can lead to more concentration of retail activities within the catchment area of the metro or the provision of a larger number of rented dwellings adjacent to stations. In fact, the Dubai Municipality has started exploring the concept of transit-oriented-development (TOD) around metro stations. In a study on the effect of proximity to a railway station on the value of residential properties in San Jose, California, Mathur and Ferrell (2013) suggest that at TOD sites the impact on the value of properties is not at all negative. Additionally, the results suggest a potential to gain acceptance and willingness from developers to connect metro stations more closely to their development instead of providing stations at the edges, as is the current case.

Assuming that the findings from this study are the average impact of the Dubai Metro on the value of properties, an attempt was made to estimate roughly the size of the uplift and the depreciation in values across Dubai. This is calculated using the average metro effect within each catchment area, as indicated in this study, multiplied by the value of all of residential and retail properties located within the catchment area of the metro (the number of properties is provided from the RTA DSTM planning data). It was found that residential properties located close to a station reduced in value by about DHS13 billion (US \$3.6), whereas the uplift in the value of residential and retail properties in the positive catchments was DHS 30 billion (US \$8.2). Given that the capital cost of the metro is DHS 30 billion, the net value increase due to the metro accounts for about 59% of its capital costs. This result implies that should Dubai decide to capture at least a part of the value increase and compensate the properties which experienced depression in values, the value capture would contribute to a relatively large percentage of the metro costs.

10.5 Contributions and implications for academic research

This study has contributed to the existing research in the following ways. First, since the empirical evidence on the impact of railways on land and property values is diverse, this research has provided the most comprehensive meta-analysis to have examined the sources of variation in estimates. This can be used as a benchmark for case studies that lack a study of railway effects on land or property values. Second, the research has objectively estimated for the first time the effect of the newly opened Dubai Metro on residential and retail property values.

Third, since a limited database was available in Dubai for records on properties, the study established four comprehensive datasets for future studies (sale transactions of dwellings, sale listings of dwellings, rental listings of dwellings and sale transactions of retail properties). The new datasets contain property values, property characteristics, building attributes, distances to amenities, distances to metro stations, distances to the nearest highway, employment density and shopping trips, both before and after the opening of the metro.

Fourth, almost all existing empirical work has considered one data structure to estimate the effect of a railway on property values. In this study, however, both repeated cross-sectional data and pseudo panel data constructed from the same dataset were used, with an extensive justification of the preferred data structure (the former). Nonetheless, the results using the latter data structure (pseudo panel data) were also reported, revealing some misleading estimates.

Fifth, although studies have indicated the need to measure the impact of the change in the travel times and costs due to a transport system on property values, to the best of our knowledge, this research represents the first attempt to study the effect of the change in the generalized cost of travel due to the introduction of a railway system on property values. This supplements the existing literature by providing an additional measure of the accessibility offered by railways.

10.6 Limitations

Although this research has, as far as possible, attempted to utilize the available transport and property data in Dubai in order to estimate the impact of the metro on property values, there are some data and methods limitations to the study. First, the study considers a sample of properties in Dubai which covers only some parts of the urban area, since no other consistent data was available for the other neighbourhoods in the emirate.

Second, the original datasets contained a limited number of property and building attributes, even though there are other variables that can affect property values, such as the number of flats in a building and the floor number where a property is located. Nevertheless, strenuous efforts were made not only to enrich the datasets as far as possible, but also to control for time-invariant unobserved heterogeneity across properties by clustering properties to location and size in the repeated cross-sectional datasets.

Third, estimates of the GC of travel for public and private transport were obtained with and without the metro, instead of actual GC of travel figures, since no other data was available. Since this value keeps all the planning and land use data constant, however, it remains possible to consider the effect of the Dubai Metro alone on the overall travel times and costs. If actual figures become available, future research can use them.

Fourth, the DID and HP methods considered in this study provide a global estimate for the effect of the Dubai Metro, without separating the effect spatially. Although there are other recently used advanced methodologies, like the GWR which allows one to estimate local effects, and the spatial DID, which allows for spatial links between properties, these are not considered here due to the relatively limited amount of spatially distributed property data.

Fifth, unobserved effects (e.g. the noise levels surrounding the property) are assumed to be time-invariant for the period of the study. Finally, due to the limited number of records per pseudo panel group, the effect is examined mainly at three catchment areas (0.5 km, 1 km and 1.5 km), although one model using repeated cross-sectional data did consider effects at finer distance bands.

10.7 Directions for future research

There is a potential for future studies to develop this research. These are divided into studies on the specific case of the Dubai Metro, and other research or case studies estimating the effect of a transport system on the value of land or property more generally.

In the case of the Dubai Metro, there are a number of studies that could derive from this research. First, since this study examined the short-term effect of the elevated part of the metro Red Line on the value of properties, there remains a room to explore the effect of the elevated and the underground parts of the metro given that more data becomes available spatially. Second, the effect of the metro on the value of properties may also be studied after the opening of the metro Green Line and after the metro system stabilizes (e.g. after 5 years of operations). It would be expected that the results would differ, especially since some differences in the effect on the sale asking values versus the sale transaction values of dwellings have already been noted here. For example, the impact radius of the metro on the sale asking values is larger than that for the transaction values. Estimating the impact after system stabilization has also been considered in previous studies (as explained in chapter 2) with varying results being obtained immediately following the opening of a railway compared to the findings a few years later (e.g. the results of Nelson and McCleskey (1989) vs. the results of Bowes and Ihlanfeldt (2001)).

Third, since this study has estimated the impact of proximity to a metro station, future research could consider the effect of proximity to a metro line as well. This can be relevant to Dubai since one of the findings in this study shows that the metro has a depressing effect on the value of dwellings located within 500 m of a metro station, due to the increased levels of noise and traffic. Fourth, given that a wider geographical distribution of property data is available in Dubai, future studies can distinguish the impact of the metro on the value of properties located at neighbourhoods with a different modal split share.

The metro could also induce changes to the travel behaviour of Dubai residents regarding the use of public transport systems, therefore a different impact of the metro on property values might be observed at a later date. Further research that considers the changes (if any) in the travel behaviour of commuters, and the preferences of users towards public and private transport services after a few years of operations, together with the resultant effect on property values may be worthwhile. Fifth, due to limited land data, this research only estimated the impact on property values but the impact on land values could be examined if historical data were collected.

Sixth, although suitable empirical methods were used for the datasets employed here (i.e. the DID and HP models), if a larger amount of spatially varying land and property data became available, GWR and SDID methods could be considered. Seventh, another dimension that could be included in future studies is the effect on property values of transport policy changes (if any) due to the operations of the Dubai Metro. Finally, since this study finds positive effects of the metro on the value of some properties, future research could consider in detail the implications for revenue raising policy.

A few suggestions are also explored for future research estimating the effect of a transport system on land and property values more generally. First, as the literature does not distinguish explicitly the ideal characteristics of control groups or reference groups versus treated groups, a more in depth study can consider this. Second, the feedback effect of the railway on land use distribution can be considered. In particular, a factor that can be considered in estimating the effect of a transport system on land and property values is the redistribution of population or employment densities due to the transport system (if any) and the feedback effect of this on land and property values.

This study has also estimated the effect on property values of the changes in the GC of travel with and without the metro. A potential avenue for future research could consider similar approaches for a new or an improved transport system, not only for railways but also bus rapid transit systems and highways. Finally, to test for the overall effect of a transport system in a city, it is suggested that consideration be given to examining the collective effect of the changes in the GC of travel and the redistribution of land uses (if any) due to the operations of or improvement in transport systems.

REFERENCES

ABADIE, A. 2005. Causal Inference. *Encyclopedia of Social Management,* 1**,** 259-266.

AGOSTINI, C. A. & PALMUCCI, G. A. 2008. The anticipated capitalisation effect of a new metro line on housing prices. *Fiscal Studies,* 29**,** 233-256.

AHLFELDT, G. M. 2013. If we build, will they pay? predicting property price effects of transport innovations. *Environment and Planning A,* 45**,** 1977-1994.

AHLFELDT, G. M. & WENDLAND, N. 2009. Looming stations: valuing transport innovations in historical context. *Economics Letters,* 105**,** 97-99.

AL-MOSAIND, M. A., DUEKER, K. J. & STRATHMAN, J. G. 1993. Light rail transit stations and property values: A hedonic price approach. *Transportation Research Record: Journal of the Transportation Research Board,* No. 1400**,** 90-94.

ALONSO, W. 1964. *Location and land use: toward a general theory of land rent,* Cambridge, Harvard University Press.

ARMSTRONG, R. J. 1994. Impacts of commuter rail service as related in single-family residential property values. *Transportation Research Record: Journal of the Transportation Research Board,* No. 1466**,** 88-98.

ASHENFELTER, O. & CARD, D. 1985. Using the longitudinal structure of earnings to estimate the effect of training programs. *The Review of Economics and Statistics,* 67**,** 648-660.

BAE, C. C., JUN, M. & PARK, H. 2003. The impact of Seoul's subway Line 5 on residential property values. *Transport Policy,* 10**,** 85-94.

BANISTER, D. 2007. Quantification of the non-transport benefits resulting from rail investment. *Transport Studies Unit - Working Paper No. 1029.* Oxford: Oxford University Centre for the Environment.

BAUM-SNOW, N. & KAHN, M. E. 2000. The effects of new public projects to expand urban rail transit. *Journal of Public Economics,* 77**,** 241-263.

BENJAMIN, J. D. & SIRMANS, G. S. 1996. Mass transportation, apartment rent and property values. *The Journal of Real Estate Research,* 12**,** 1-8.

BERTRAND, M., DUFLO, E. & MULLAINATHAN, S. 2004. How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics,* 119**,** 249-275.

BILLINGS, S. B. 2011. Estimating the value of a new transit option. *Regional Science and Urban Economics,* 41**,** 525-536.

BOLLINGER, C. R., IHLANFELDT, K. R. & BOWES, D. R. 1998. Spatial variation in office rents within the Atlanta Region. *Urban Studies,* 35**,** 1097-1118.

BOWES, D. R. & IHLANFELDT, K. R. 2001. Identifying the impacts of rail transit stations on residential property values. *Journal of Urban Economics,* 50**,** 1-25.

BRINCKERHOFF, P. 2001. The effect of rail transit on property values: a summary of studies. Neorail II Cleveland, Ohio.

BRUNSDON, C., FOTHERINGHAM, A. S. & CHARLTON, M. 1998. Geographically weighted regression - modelling spatial non-stationarity. *The Statistician,* 47**,** 431-443.

CAMERON, A. C. & TRIVEDI, P. K. 2005. *Microeconometrics: Methods and applications,* New York, Cambridge University Press.

CERVERO, R. 2003. Effects of light and commuter rail transit on land prices: experiences in San Deigo County. University of California.

CERVERO, R. & DUNCAN, M. 2002. Transit's value-added: effects of light and commuter rail services on commercial land values. *Transportation Research Record: Journal of the Transportation Research Board,* No. 1805**,** 8-15.

CERVERO, R. & SUSANTONO, B. 1999. Rent capitalization and transportation infrastructure development in Jakarta. *Review of Urban and Regional Development Studies,* 11**,** 11-23.

CHEN, H., RUFOLO, A. & DUEKER, K. J. 1997. Measuring the impact of light rail systems on single family home values: A hedonic approach with GIS application. Report Centre for Urban Studies, Portland State University.

CHESTERTON 2000. Property Market Study - working paper no.32.

CLARK, T. S. & LINZER, D. A. 2013. Should I use fixed or random effects? Available: http://userwww.service.emory.edu/~tclark7/randomeffects.pdf.

CLOWER, T. L. & WEINSTEIN, B. L. 2002. The impact of Dallas (Texas) area rapid transit light rail stations on taxable property valuations. *Australasian Journal of Regional Studies,* 8**,** 389-400.

COLLADO, M. D. 1997. Estimating dynamic models from time series of independent cross-sections. *Journal of Econometrics,* 82**,** 37-62.

CONCAS, S. 2012. Accessibility and housing price resilience: evidence from limited-access roadways. Center for Urban Transportation Research, College of Engineering, University of South Florida.

COOPER, P. J. 2001. Available: http://www.ameinfo.com/16342.html [Accessed 26-04-2011 2011].

DEATON, A. 1985. Panel data from time series of cross-sections. *Journal of Econometrics,* 30**,** 109-126.

DEBREZION, G., PELS, E. & RIETVELD, P. 2006. The impact of rail transport on real estate prices: an empirical analysis of the Dutch housing markets. Tinbergen Institute Discussion Paper.

DEBREZION, G., PELS, E. & RIETVELD, P. 2007. The impact of railway stations on residential and commercial property value: A meta-analysis. *Journal of Real Estate Finance and Economics,* 35**,** 161-180.

DECOSTER, J. 2004. Meta-analysis notes. Department of Psychology University of Alabama.

DEWEES, D. N. 1976. The effect of a subway on residential property values in Toronto. *Journal of Urban Economics,* 3**,** 357-369.

DIAZ, R. B. Impacts of rail transit on property values. *Commuter Rail/Rapid Transit Conference, Proceedings*, 1999 Toronto. American Public Transit Association, 66-73.

DICKENS, I. 1992. Transport investment, economic development and strategic planning: The example of light rail transit *Planning Practice and Research,* 7**,** 9-12.

DU, H. & MULLEY, C. 2006. Relationship between transport accessibility and land value. Local model approach with geographically weighted regression. *Transportation Research Record: Journal of the Transportation Research Board,* No. 1977**,** 197-205.

DU, H. & MULLEY, C. 2007. Transport accessibility and land values: a case study of Tyne and Wear. *RICS Research paper series.*

DUBAI MUNICIPALITY 2003. A study on alternative systems for public transport in Dubai: PS002. *In:* DUBAI MUNICIPALITY (ed.).

DUBAI MUNICIPALITY 2004. Ridership forecast report for project PS007 preliminary engineering stage for Dubai rail project. *In:* MUNICIPALITY, D. (ed.).

DUBAI MUNICIPALITY 2005. Project PS007 preliminary engineering stage for Dubai rail project. *In:* MUNICIPALITY, D. (ed.).

DUBAI MUNICIPALITY 2010. Dubai urban master plan 2020: phase 1 - synthesis report. *In:* MUNICIPALITY, D. (ed.).

DUBAI STATISTICS CENTER 2002. Statistical year book - Emirate of Dubai 2002. http://dsc.gov.ae/Publication/SYB02.PDF.

DUBAI STATISTICS CENTER. 2010. Dubai in figures 2010. Available: http://dsc.gov.ae/EN/Publications/Pages/PublicationsList.aspx?PublicationId=2&Year=2010.

DUBAI STATISTICS CENTER. 2011a. Dubai in figures. Available: http://dsc.gov.ae/EN/Publications/Pages/PublicationsList.aspx?PublicationId=2&Year=2011.

DUBAI STATISTICS CENTER 2011b. Statistical yearbook - Emirate of Dubai 2011. http://dsc.gov.ae/Publication/Statistical%20Year%20Book%202011.pdf.

DUBAI STATISTICS CENTER. 2012. Bulletin of labour fource survey results 2011. Available: http://dsc.gov.ae/EN/Publications/Pages/PublicationsList.aspx?PublicationId=4&Year=2011.

DUBÉ, J., LEGROS, D., THÉRIAULT, M. & DES ROSIERS, F. 2014. A spatial Difference-in-Differences estimator to evaluate the effect of change in public mass transit systems on house prices. *Transportation Research Part B: Methodological,* 64**,** 24-40.

DUBÉ, J., THÉRIAULT, M. & DES ROSIERS, F. 2013. Commuter rail accessibility and house values: The case of the Montreal South Shore, Canada, 1992–2009. *Transportation Research Part A: Policy and Practice,* 54**,** 49-66.

DUNCAN, M. 2008. Comparing rail transit capitalization benefits for single-family and condominium units in San Diego, California. *Transportation Research Record: Journal of the Transportation Research Board,* No. 2067**,** 120-130.

EDWARDS, A. & COOPER, M. 2009. The impact of the Dubai Metro on real estate values. Dubai: DTZ.

EFTHYMIOU, D. & ANTONIOU, C. 2013. How do transport infrastructure and policies affect house prices and rents? Evidence from Athens, Greece. *Transportation Research Part A: Policy and Practice,* 52**,** 1-22.

ENOCH, M. 2002. Recouping public transport costs from gains in land values. *Traffic Engineering and Control,* 43**,** 336-340.

ENOCH, M., POTTER, S. & ISON, P. 2005. A strategic approach to financing public transport through property values *Public Money & Management***,** 147-154.

FOTHERINGHAM, A. S., BRUNSDON, C. & CHARLTON, M. 2002. *Geographically Weighted Regression*, Wiley.

FRANK, K. A. 2000. Impact of a confounding variable on a regression coefficient. *Sociological methods & research,* 29**,** 147-194.

FTA 2000. Transit benefits 2000 working papers - A public choice policy analysis. *In:* U.S. DEPARTMENT OF TRANSPORTATION (ed.). Washington D.C.

FUJITA, M. 1989. *Urban economic theory: land use and city cize,* Cambridge MA, Cambridge University Press.

GARGIULO, C. & DE CIUTIS, F. 2010. Urban transformation and property value variation. *Journal of Mobility, Land Use and Environment,* 3**,** 65-84.

GATZLAFF, D. H. & SMITH, M. T. 1993. The impact of the Miami Metrorail on the value of residences near station locations. *Land Economics,* 69**,** 54-66.

GIBBONS, S., LYYTIKAINEN, T., OVERMAN, H., SANCHIS-GUARNER, R. & LAIRD, J. 2010. Evaluating the productivity impacts of road transport schemes - report on pilot study findings

GIBBONS, S. & MACHIN, S. 2005. Valuing rail access using transport innovations. *Journal of Urban Economics,* 57**,** 148-169.

HAHN, J., HAM, J. & MOON, H. R. 2011. Test of random versus fixed effects with small within variation. *Economics Letters,* 112**,** 293-297.

HSIAO, C. 2003. *Analysis of Panel Data*, Cambridge University Press.

HUI, E. C. & HO, V. S. 2004. Land value capture mechanisms in Hong Kong and Singapore. A comparative analysis. *Journal of Property Investment & Finance,* 22**,** 76-100.

ISARD, W. & PECK, M. J. 1954. Location theory and international and interregional trade theory. *The Quarterly Journal of Economics,* 68**,** 97-114.

KIM, J. & ZHANG, M. 2005. Determining transit's impact on Seoul commercial land values: an application of spatial econometrics. *International Real Estate Review,* 8**,** 1-26.

KOBAYASHI, T. & LANE, B. 2007. Spatial heterogeneity and transit use. *11th World Conference on Transportation Research.* Berkeley California USA.

KOSTER, H. R. A., OMMEREN, J. N. V. & RIETVELD, P. 2010. Estimating the benefits of improved rail access: geographical range and anticipation effects. *Tinbergen Institute Discussion Paper.* Tinbergen Institute.

KRUGMAN, P. R. 1993. On the relationship between trade theory and location theory. *Review of International Economics,* 1**,** 110-122.

LAAKSO, S. 1992. Public transport investment and residential property values in Helsinki. *Scandinavian Housing and Planning Research,* 9**,** 217-229.

LANDIS, J., GUHATHAKURTA, S. & ZHANG, M. 1994. Capitalization of transit investments into single-family home prices: a comparative analysis of five California Rail Transit Systems.

LANE, R., POWELL, T., EYERS, T., PARIS, J., LUCAS, K. & JONES, P. 2004. JLE summary report. University of Westminister: Transport for London and the Department for Transport.

LARI, A., LEVINSON, D., ZHAO, Z., IACONO, M., AULTMAN, S., DAS, K. V., JUNGE, J., LARSON, K. & SCHARENBROICH, M. 2009. Value capture for transportation finance: technical research report. *Technical report CTS 09-18.* Centre for transportation studies: University of Minnesota.

MARTINEZ, L. M. & VIEGAS, J. M. 2009. Effects of transportation accessibility on residential property values - hedonic price models in the Lisbon, Portugal, ,metropolitan area. *Transportation Research Record: Journal of the Transportation Research Board,* No. 2115**,** 127-137.

MARTÍNEZ, L. M. & VIEGAS, J. M. 2012. The value capture potential of the Lisbon Subway. *Journal of Transport and Land Use,* 5.

MATHUR, S. & FERRELL, C. 2013. Measuring the impact of sub-urban transit-oriented developments on single-family home values. *Transportation Research Part A: Policy and Practice,* 47**,** 42-55.

MCCANN, P. 2001. *Urban and regional economics*, Oxford University Press.

MCDONALD, J. & OSUJI, C. 1995. The effect of anticipated transportation improvement on residential land values. *Regional Science and Urban Economics,* 25**,** 261.

MCMILLEN, D. P. & MCDONALD, J. 2004. Reaction of house prices to a new rapid transit line: Chicago's midway line, 1983-1999. *Real Estate Economics,* 32**,** 463-486.

MEDDA, F. 2008. Land value mechanisms in transport PPPs: investment in urban mass transit systems. *World Bank Publications (2008 Forthcoming).*

MEDDA, F. 2012. Land value capture finance for transport accessibility: a review. *Journal of Transport Geography,* 25**,** 154-161.
MELO, P. C., GRAHAM, D. J. & NOLAND, R. B. 2009. A meta-analysis of estimates of urban agglomeration economies. *Regional Science and Urban Economics,* 39**,** 332-342.
MENNIS, J. 2006. Mapping the results of goegraphically weighted regression. *The Cartographic Journal,* 43**,** 171-179.
MIKELBANK, B. A. 2004. Spatial analysis of the relationship between housing values and investments in transportation infrastructure. *The Annals of Regional Science,* 38**,** 705-726.
MOHAMMAD, S. I., GRAHAM, D. J., MELO, P. C. & ANDERSON, R. J. 2013. A meta-analysis of the impact of rail projects on land and property values. *Transportation Research Part A: Policy and Practice,* 50**,** 158-170.
MUTH, R. F. 1969. *Cities and housing; the spatial pattern of urban residential land use,* Chicago, University of Chicago Press.
MVA 2011. Task 5: Develop a methodology for assessment of enhanced land value due to transportation infrastructure improvements. *In:* ROADS AND TRANSPORT AUTHORITY (ed.). Dubai.
NELSON, A. C. & MCCLESKEY 1989. Influence of elevated transit stations on neighborhood house values. Georgia Institute of Technology.
O'SULLIVAN, A. 2003. *Urban Economics*, McGraw-Hill Higher Education.
PAN, H. & ZHANG, M. 2008. Rail transit impacts on land use: evidence from Shanghai, China. *Transportation Research Record: Journal of the Transportation Research Board,* No. 2048**,** 16-25.
PEARL, J. 2009. Causal inference in statistics: An overview. 96-146.
PLÜMPER, T. & TROEGER, V. E. 2007. Efficient estimation of time-invariant and rarely changing variables in finite sample panel analyses with unit fixed effects. *Political Analysis,* 15**,** 124-139.
PREDÖHL, A. 1928. The theory of location in its relation to general economics. *Journal of Political Economy,* 36**,** 371-390.
RICS 2004. Land value and public transport. Stage two - summary of findings. Department for Transport, Office of the Deputy Prime Minister.
RICS POLICY UNIT 2002. Land value and public transport: Stage 1 - Summary of findings. Office of the Deputy Prime Minister.
ROADS AND TRANSPORT AUTHORITY 2010a. Dubai Strategic Transport Model. Strategic Planning Department.
ROADS AND TRANSPORT AUTHORITY 2010b. The resutls of the survey on needs of the Dubai Metro users. Development and Corporate Performance Department.
ROADS AND TRANSPORT AUTHORITY 2011. The resutls of the survey on needs of the Dubai Metro users. Development and Corporate Performance Department.
ROADS AND TRANSPORT AUTHORITY 2012a. Development of public transport system in Dubai. Strategic Planning Department.
ROADS AND TRANSPORT AUTHORITY 2012b. RTA statistics 2011.
ROSEN, S. 1974. Hedonic prices and implicit markets: product differentiation in pure competition. *The Journal of Political Economy,* 28**,** 34-55.
RYAN, S. 1999. Property values and transportation facilities: finding the transportation-land use connection. *Journal of Planning Literature,* 13**,** 412-427.

SALON, D. & SHEWMAKE, S. 2012. Opportunities for value capture to fund public transport: A comprehensive review of the literature with examples from East Asia. *Transportation Research Board.* Washington, DC. USA: TRB 2012 Annual Meeting.

SAPHORES, J. D. & YEH, C. C. 2013. Time on market and the impact of a mass rapid transit system on residential property values - The case of Kaohsiung, Taiwan. *Transportation Research Board.* Washington, DC. USA: TRB 2013 Annual Meeting.

SEO, K., GOLUB, A. & KUBY, M. 2014. Combined impacts of highways and light rail transit on residential property values: a spatial hedonic price model for Phoenix, Arizona. *Journal of Transport Geography,* 41**,** 53-62.

SMITH, J. & GIHRING, T. 2006. Financing transit systems through value capture. *The American journal of economics and sociology,* 65**,** 751-786.

SMITH, J. J., GIHRING, T. A. & LITMAN, T. 2010. Financing transit systems through value capture - an annotated bibliography. *Victoria Transport Policy Institute.*

STANLEY, T. D. & JARRELL, S. B. 2006. Meta-regression analysis: a quantitative method of literature surveys. *Journal of Economic Surveys,* 3**,** 161-170.

TSAI, C., LEONG, W., MULLEY, C. & CLIFTON, G. 2013. Examining estimation bias and efficiency for pseudo panel data in travel demand analysis. *Transportation Research Record: Journal of the Transportation Research Board,* No. 2354**,** 1-8.

VERBEEK, M. & NIJMAN, T. 1992. Can cohort data be treated as genuine panel data. *Empirical Economics,* 17**,** 9-23.

VICHIENSAN, V. & MIYAMOTO, K. 2010. Influence of urban rail transit on house value: spatial hedonic analysis in Bangkok. *Journal of the Eastern Asia Society for Transportation Studies,* 8.

VOITH, R. 1991. Transportation, sorting and house values. *Real Estate Economics,* 19**,** 117-137.

WEICHSELBAUMER, D. & WINTER-EBMER, R. 2005. A meta-analysis of the international gender wage gap. *Journal of Economic Surveys,* 19**,** 479-511.

WEINBERGER, R. R. 2001. Light rail proximity: benefit or detriment?: The case of Santa Clara County, California. *Transportation Research Record: Journal of the Transportation Research Board,* No. 1747**,** 104-113.

WEINSTEIN, B. L. & CLOWER, T. L. 1999. The initial economic impacts of the DART LRT system. University of North Texas: Centre for Economic Development and Reseach.

WEIS, C. & AXHAUSEN, K. W. 2009. Induced travel demand: Evidence from a pseudo panel data based structural equations model. *Research in Transportation Economics,* 28**,** 8-18.

WU, W. 2012. Does public investment spur the land market?: Evidence from transport improvement in Beijing. *SERC Discussion paper 116.* Department of Geography and Environment, London School of Economics and Polical Sciences.

YANKAYA, U. & CELIK, H. M. 2004. Modeling the impact of rail transit investment on the values of residential property: A hedonic price approach in the case of Izmir Subway, Turkey. Izmir Institute of Technology: Department of Civil Engineering.

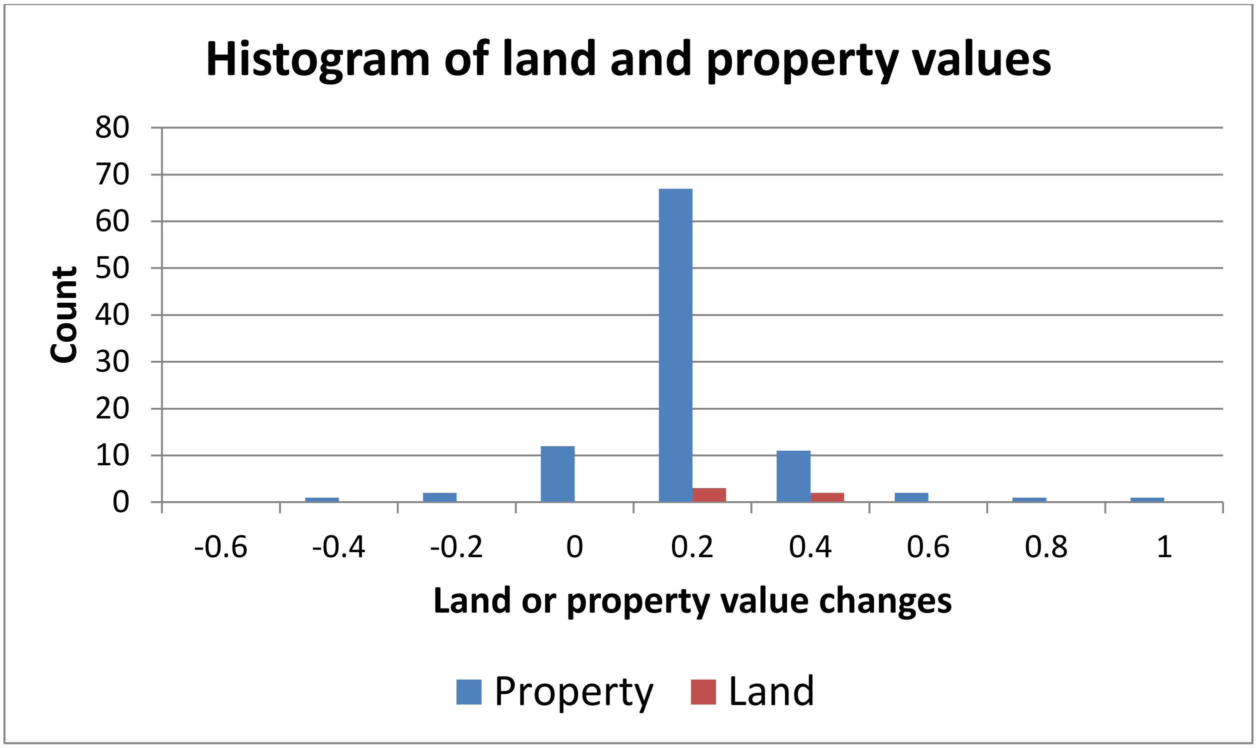
ZHANG, M. 2009. Bus verus rail: Meta-analysis of cost characteristis, carrying capacities, and land use impacts. *Transportation Research Record: Journal of the Transportation Research Board,* No. 2110**,** 87-95.

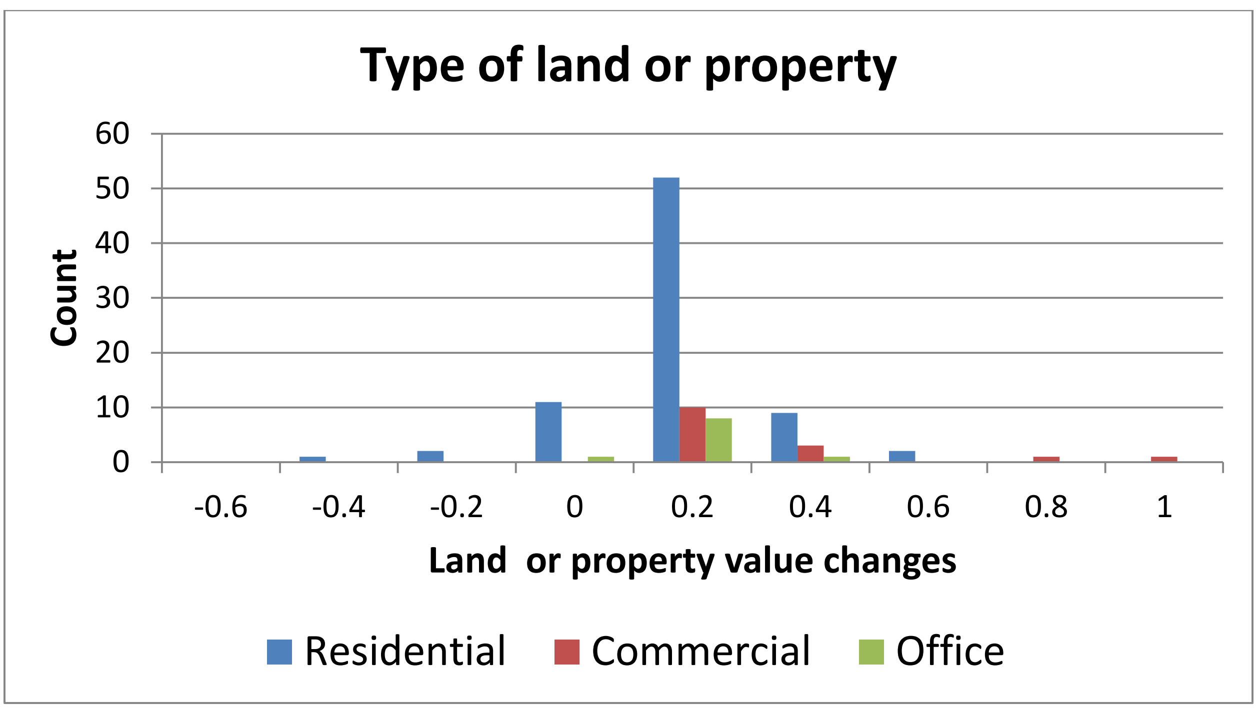
ZHANG, M. & WANG, L. 2013. The impacts of mass transit on land development in China: the case of Beijing. *Research in Transportation Economics,* 40**,** 124-133.

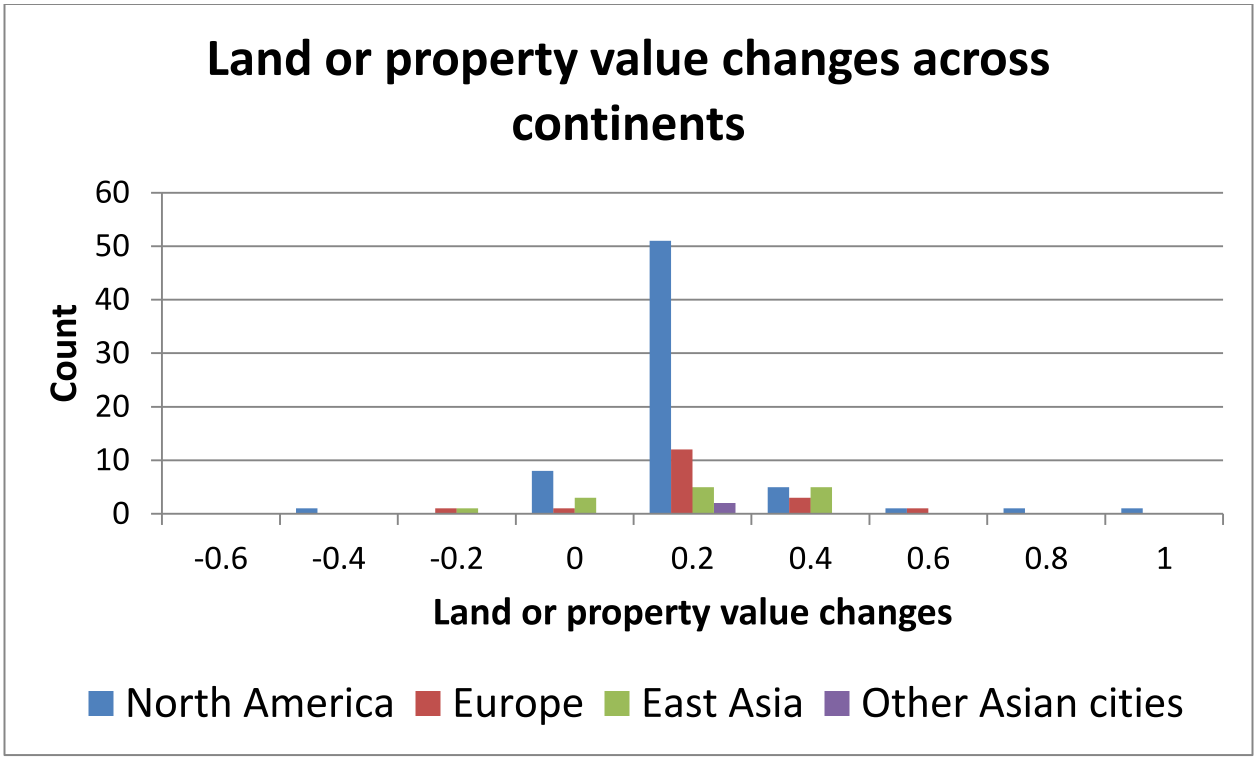
ZHOU, Y. & ZHANG, J. Spatial data analysis of Beijing-Shanghai (B-S) railway based on GWR and OLS model. Third International Conference on Information and Computing, 2010 Wuxi, Jiang Su. 125-128.

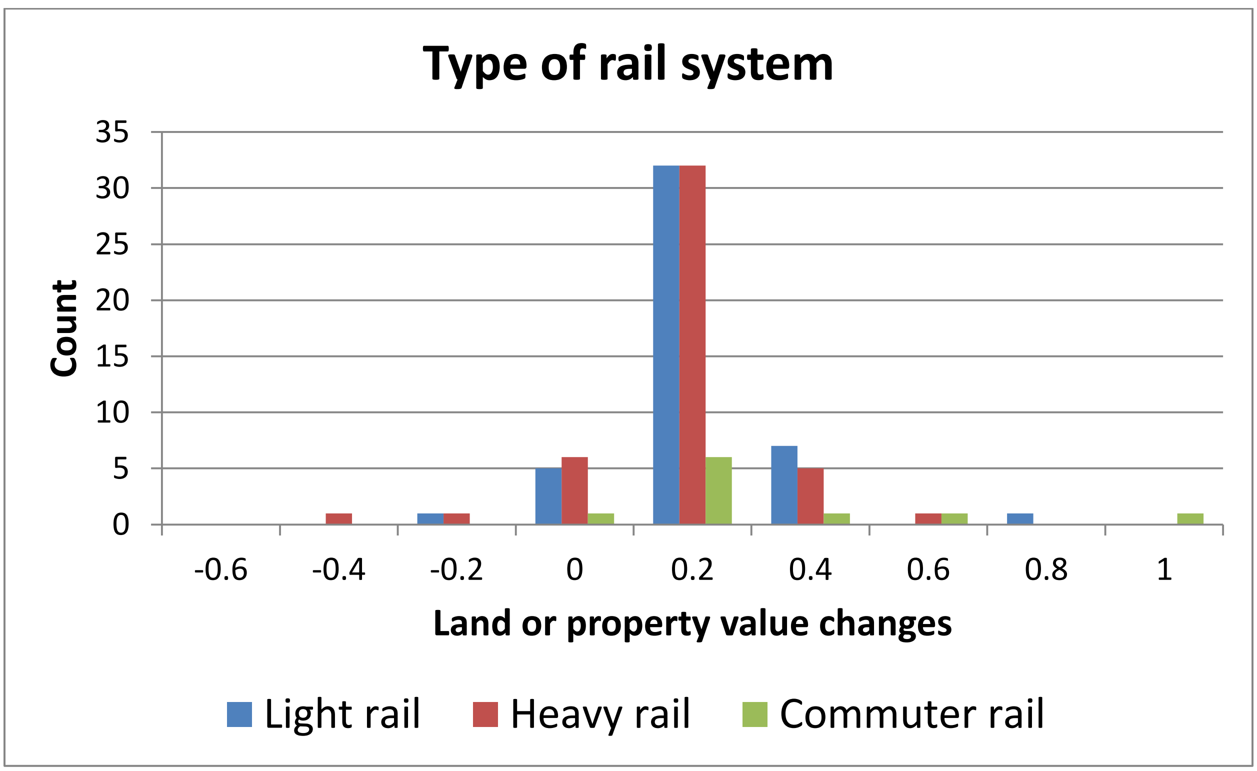
Chapter 11. APPEDNICES

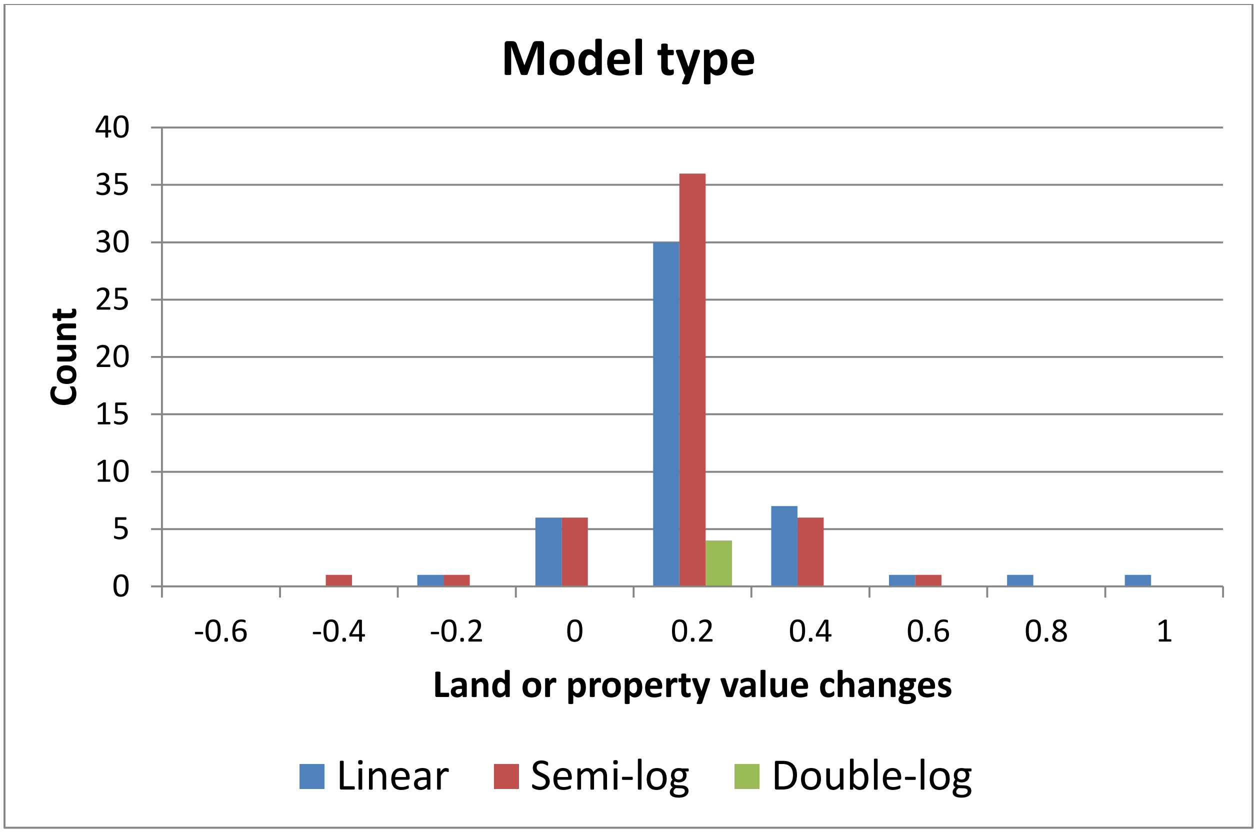
Appendix A – Histograms of the meta-sample






Land or property value changes across continents
Count
60
50
40
30
20
10
0
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1
Land or property value changes
North America
Europe
East Asia
Other Asian cities


Type of rail system
Count
35
30
25
20
15
10
5
0
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1
Land or property value changes
Light rail
Heavy rail
Commuter rail


Model type
40
35
30
25
20
15
10
5
0
Count
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1
Land or property value changes
Linear
Semi-log
Double-log

Appendix B – Within- and across-cohorts variation for grouping options 3 and 4

Table 11-1: Within- and between-cohort variation for grouping options 3 and 4 using sale transactions of residential properties (RERA)

Grouping observations to option 3			
Variable		**Mean**	**Std. Dev.**
ln(value)	overall	13.867	0.744
	between		0.451
	within		0.489
Area	overall	141.175	97.226
	between		89.352
	within		51.145

Grouping observations to option 4			
Variable		**Mean**	**Std. Dev.**
ln(value)	overall	13.867	0.744
	between		0.555
	within		0.379
Area	overall	141.175	97.226
	between		112.724
	within		18.743

Table 11-2: Within- and between-cohort variation for grouping options 3 and 4 using sale listings of residential properties (REIDIN)

Grouping observations to option 3			
Variable		**Mean**	**Std. Dev.**
ln(value)	overall	13.089	0.666
	between		0.647
	within		0.375
Area	overall	194.951	156.551
	between		95.821
	within		93.613
No. bedrooms	overall	2.323	1.246
	between		0.949
	within		0.693

Grouping observations to option 4			
Variable		**Mean**	**Std. Dev.**
ln(value)	overall	13.089	0.666
	between		0.671
	within		0.232
Area	overall	194.951	156.551
	between		126.731
	within		71.766
No. bedrooms	overall	2.323	1.246
	between		1.209
	within		0.000

Table 11-3: Within- and between-cohort variation for grouping options 3 and 4 using rent listings of residential properties (REIDIN)

Grouping observations to option 3			
Variable		**Mean**	**Std. Dev.**
ln(value)	overall	7.813	0.590
	between		0.635
	within		0.258
Area	overall	153.712	108.305
	between		82.900
	within		52.252
No. bedrooms	overall	1.818	1.245
	between		1.190
	within		0.000

Grouping observations to option 4			
Variable		**Mean**	**Std. Dev.**
ln(value)	overall	7.813	0.590
	between		0.635
	within		0.258
Area	overall	153.712	108.305
	between		83.898
	within		47.297
No. bedrooms	overall	1.818	1.245
	between		1.200
	within		0.000

Table 11-4: Within- and between-cohort variation for grouping options 3 and 4 using sale transactions of retail properties (RERA)

Grouping observations to option 3			
Variable		**Mean**	**Std. Dev.**
ln(value)	overall	13.340	0.879
	between		0.796
	within		0.523
Area	overall	94.848	62.387
	between		87.286
	within		31.992

Grouping observations to option 4			
Variable		**Mean**	**Std. Dev.**
ln(value)	overall	13.340	0.879
	between		1.434
	within		0.550
Area	overall	94.848	62.387
	between		336.241
	within		11.921

Appendix C – Transport data

Table 11-5: The different segments of population in Dubai (source: Dubai Municipality, 2004)

Demand segment	Trip maker group	Car availability
1	Emirati male	Car available
2	Emirati female	Car available
3	Expatriate male	Car available
4	Expatriate female	Car available
5	Expatriate male	Non car available
6	Expatriate female	Non car available
7	Emirati pupil 1 (primary school)	Partly car available
8	Emirati pupil 2 (secondary school)	Partly car available
9	Expatriate pupil 1 (primary school)	Partly car available
10	Expatriate pupil 2 (secondary school)	Partly car available
11	Students	Car available
12	Retired	Car available
13	Tourist	Non car available
14	Labourers	Non car available

Appendix D – More details on property data

Table 11-6: Number of property records in each dataset at smaller catchment areas

Catchment area	0-300m	301m to 500m	501m to 700m	701m to 900m	901m to 1,100m	1,101m to 1,300m	1,301m to 1,500m	>1,500m
	Number of observations per dataset							
	Sale transactions of residential properties							
Repeated cross-sectional	2,419	5,492	3,777	6,487	4,734	3,677	2,373	10,349
Pseudo Panel	105	118	93	171	155	151	98	2,453
	Sale listings of residential properties							
Repeated cross-sectional	9,938	14,693	37,173	31,092	13,513	8,455	1,964	49,150
Pseudo Panel	237	277	295	456	189	189	31	614
	Rent listings of residential properties							
Repeated cross-sectional	1,894	5,367	28,365	6,287	3,802	3,939	2,000	29,630
Pseudo Panel	47	107	88	92	70	66	23	330
	Sale transactions of retail properties							
Repeated cross-sectional	357	351	231	227	126	157	61	1,909
Pseudo Panel	33	42	32	33	59	55	26	56

Table 11-7 presents the number of observed properties per community in each dataset compared to the actual number of properties within that community. It is worth mentioning, however, that the actual number of properties in a community is the total number of properties, regardless of whether they are available for sale or rent or otherwise. This method, therefore, probably over-represents the actual number of available properties for sale and the number of properties for rent. As a result, the percentage of the sample size to the actual size is likely under-estimated.

The table indicates that RERA datasets contain fewer records of property transactions per community compared to that in REIDIN datasets. Our sample contains on average between 5% and 15% of the property records per community in the transactions datasets (RERA) and around 250% of that in the listings datasets (REIDIN). The fact that the REIDIN datasets contain on average 2.5 times more listing records than the actual number of properties per community implies that the same property may be listed for sale or rent more than once in a year, which also indicates that REIDIN dataset can be a panel dataset. Nevertheless, since REIDIN does not provide an ID for each property, it is not possible to trace properties.

Table 11-7: Number of property records in each dataset compared to the actual number of properties in each community (source: self-produced table based on information provided by RERA and DSC)

		No. properties		% of observed to actual	Average %
Dataset	Community	Actual	Observed		
Sale transaction of residential properties - RERA	Al Hebiah First	3,435	1,024	29.8%	14.7%
	Al Thanyah fifth	11,682	895	7.7%	
	Al Thanyah fourth	4,944	608	12.3%	
	Al Thanyah third	6,143	1,405	22.9%	
	Business Bay	13,604	1,440	10.6%	
	Dubai Marina	21,860	4,260	19.5%	
	Wadi Al Safa 6	1,949	276	14.2%	
	Warsan First	23,148	192	0.8%	
Sale listings of residential properties - REIDIN	Al Hebiah fourth	1,653	2,542	153.8%	248.2%
	Al Thanyah third	6,143	25,233	410.8%	
	Arabian Ranches	1,949	8,547	438.5%	
	Business Bay	13,604	31,875	234.3%	
	DIFC	379	2,226	587.3%	
	Dubai Marina	21,860	59,983	274.4%	
	Jabal Ali	15,617	304	1.9%	
	Jebel Ali First	7,011	1,559	22.2%	
	Palm Jumeirah	4,830	27,489	569.1%	
	Trade Centre	3,418	820	24.0%	
	Warsan First	23,148	3,140	13.6%	
Rent listings of residential properties - REIDIN	Al Thanyah fourth	4,945	15,931	322.2%	249.7%
	Al Thanyah third	6,143	17,298	281.6%	
	Arabian Ranches	1,949	11,515	590.8%	
	Business Bay	13,604	20,277	149.1%	
	DIFC	379	2,204	581.5%	
	Dubai Marina	21,860	45,433	207.8%	
	Jebel Ali First	7,011	2,773	39.6%	
	Nadd Hessa	4,188	856	20.4%	
	Warsan First	23,148	12,649	54.6%	
Sale transactions of retail properties - RERA	Al Thanyah first	706	71	10.1%	4.7%
	Al Thanyah third	494	52	10.5%	
	Al Thanyah fifth	6,301	171	2.7%	
	Business Bay	4,826	71	1.5%	
	Dubai Marina	339	10	2.9%	
	Nadd Hessa	1,063	9	0.8%	
	Warsan First	4,239	197	4.6%	

Figure 11-1: The distribution of cohort sizes for sale transactions of residential properties (RERA)

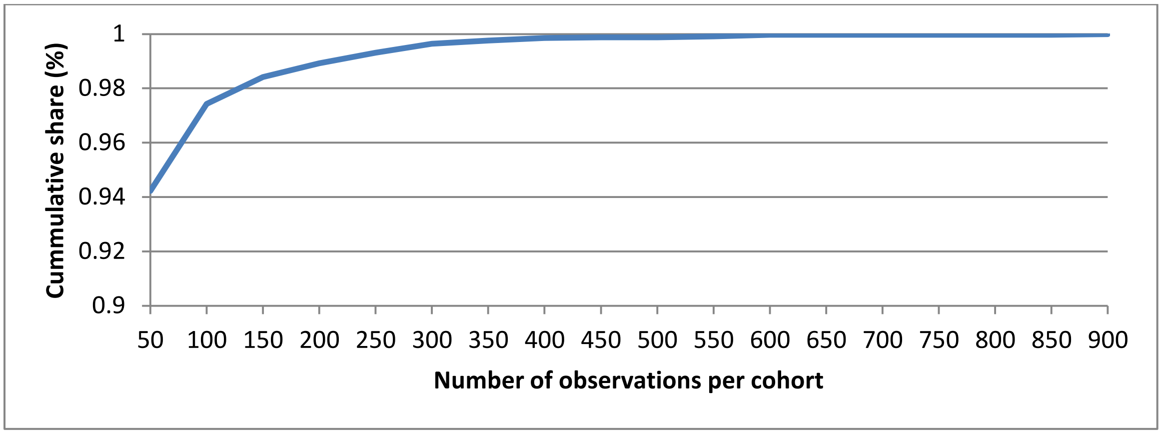


Figure 11-2: The distribution of cohort sizes for sale listings of residential properties (REIDIN)

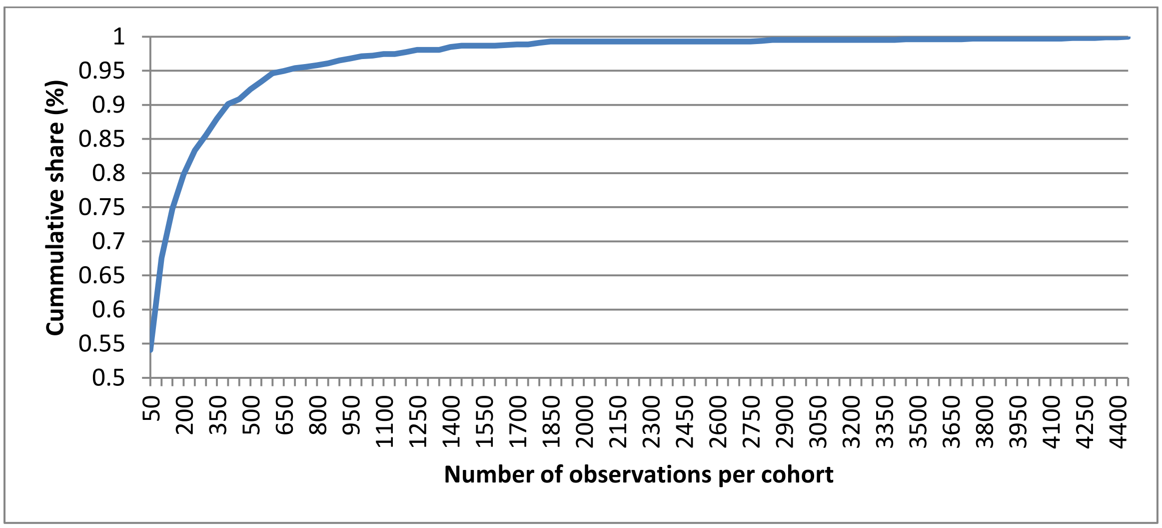


Figure 11-3: The distribution of cohort sizes for rent listings of residential properties (REIDIN)

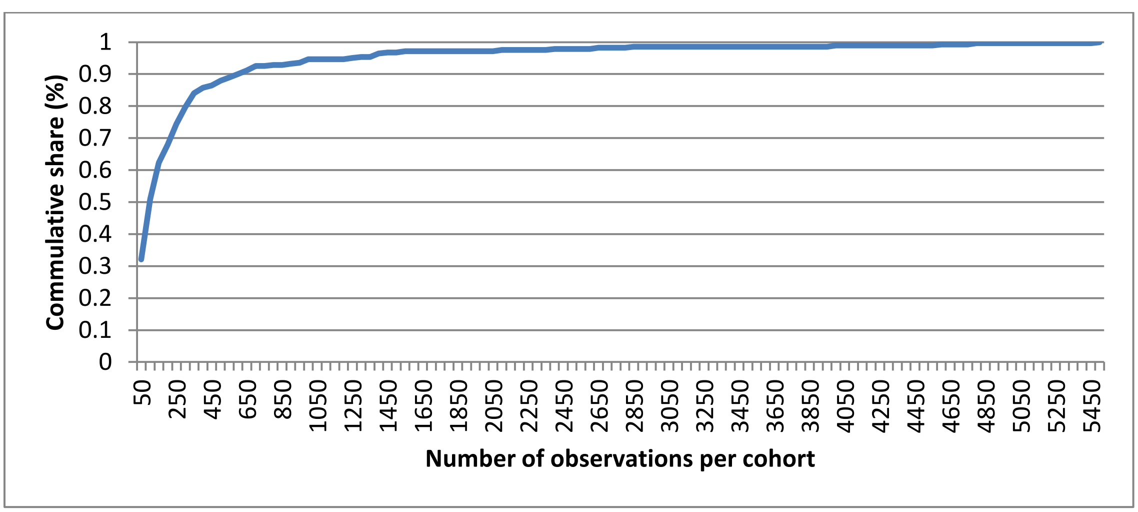
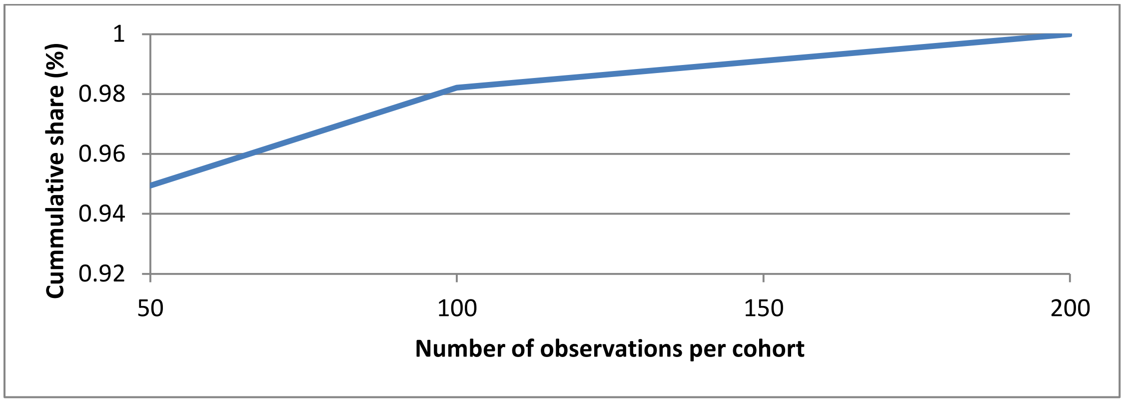
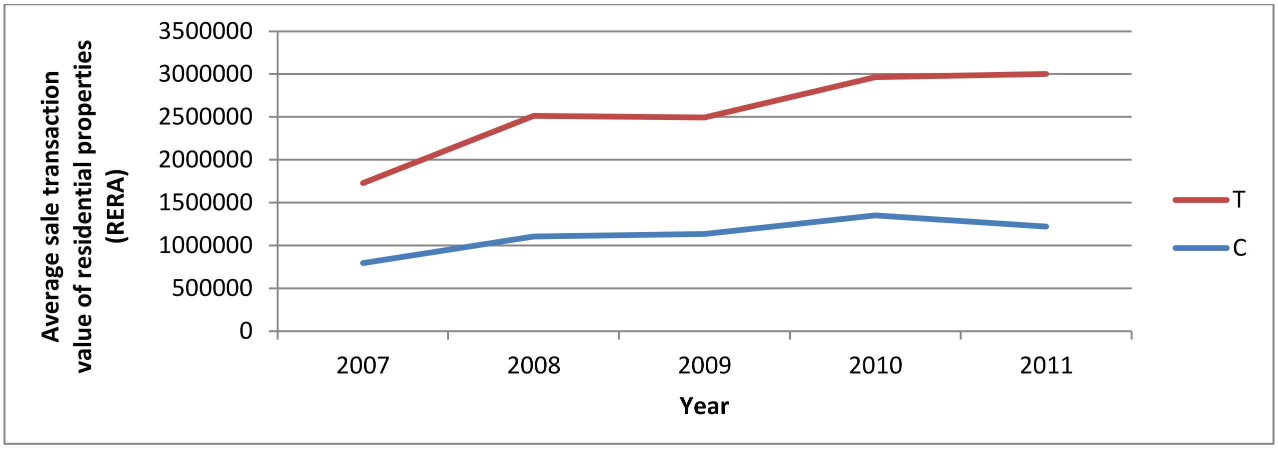


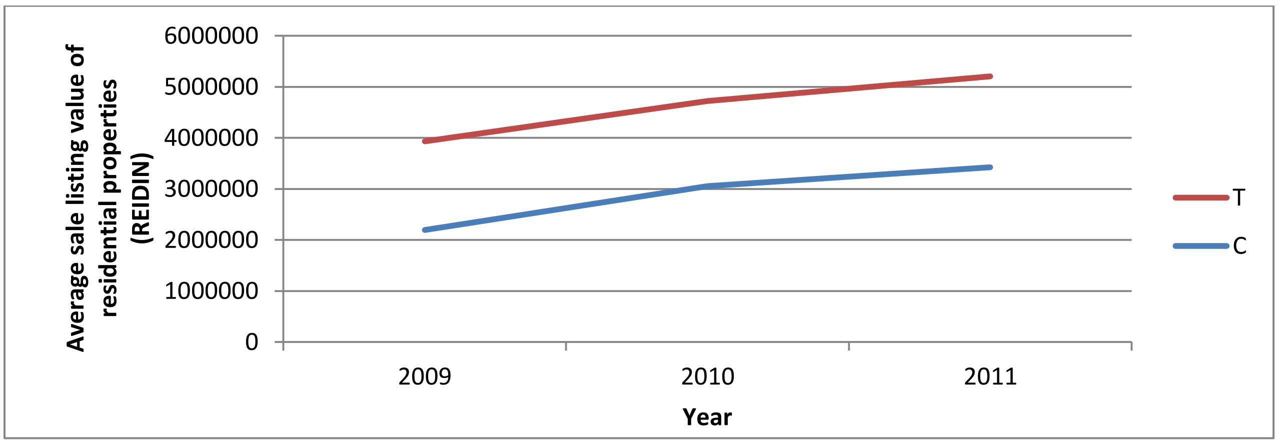
Figure 11-4: The distribution of cohort sizes for sale transactions of retail properties (RERA)

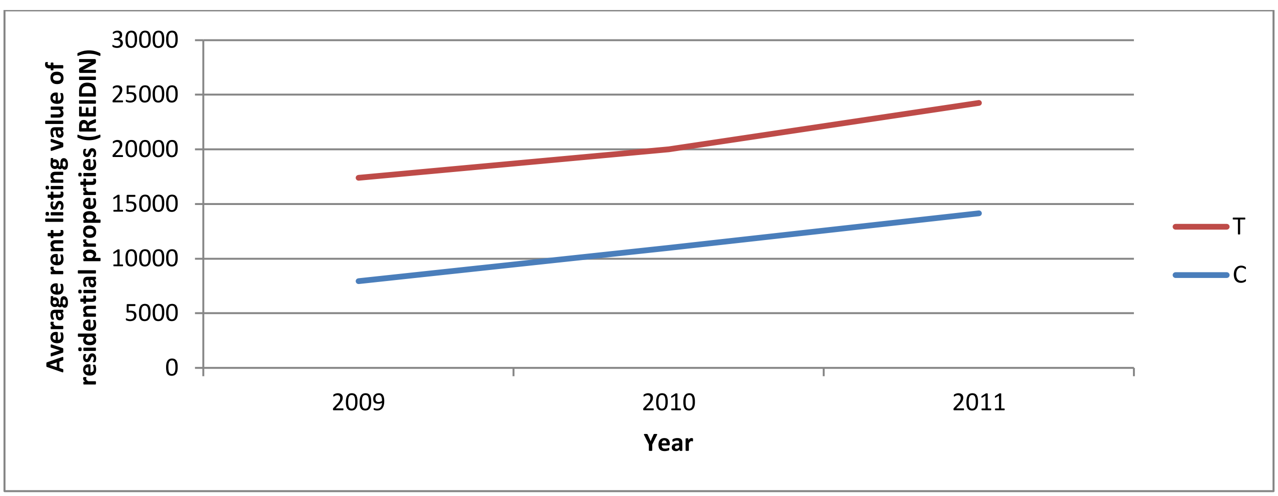


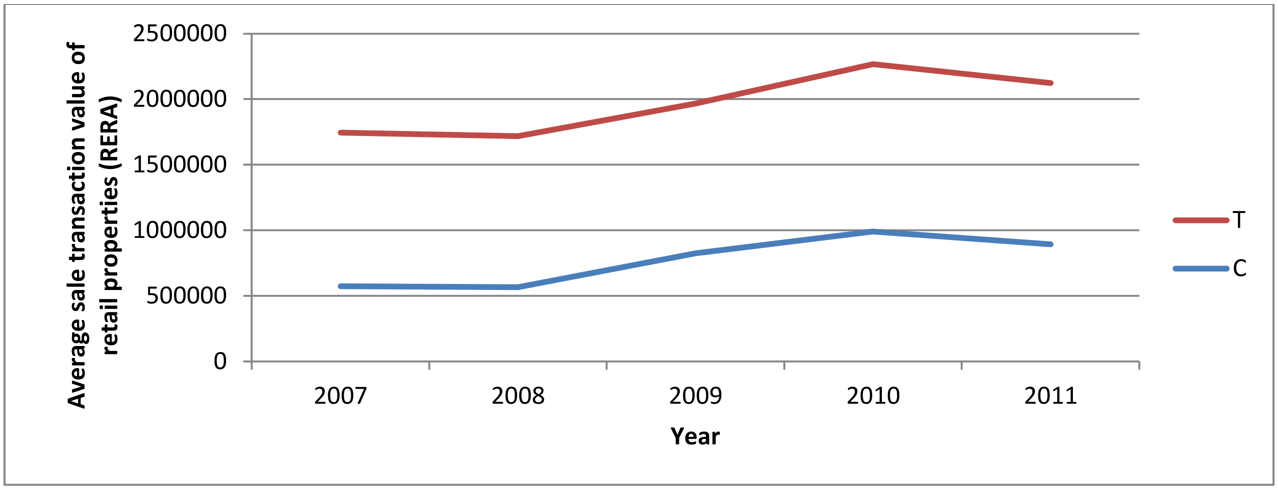
Appendix E – DID assumptions

Figure 11-5: Parallel trend assumption for the DID models (T is treated and C is control)








Average sale transaction value of retail properties (RERA)
2500000
2000000
1500000
1000000
500000
0
2007
2008
2009
2010
2011
Year
T
C

Appendix F – Main results using an FE estimator

Table 11-8: Key results for the effect of the Dubai Metro on property values using an FE estimator

Model type	**Full DID model**					
Equation number	**17**	**18**	**17**	**18**	**17**	**18**
Catchment zone	**0.5 km**		**1 km**		**1.5 km**	
Data structure	**RCS**	**PP**	**RCS**	**PP**	**RCS**	**PP**
Dataset	Sale transactions of residential properties (RERA)					
Metro effect	-0.106*** (0.014)	-0.179*** (0.064)	0.075*** (0.008)	-0.112** (0.053)	0.034*** (0.007)	-0.158*** (0.05)
R^2 *within*	0.3245	0.2020	0.3334	0.2054	0.3675	0.1895
Dataset	Sale listings of residential properties (REIDIN)					
Metro effect	-0.013** (0.006)	-0.015 (0.021)	0.033*** (0.004)	0.009 (0.017)	0.048*** (0.005)	0.050*** (0.020)
R^2 *within*	0.1250	0.4298	0.1253	0.4297	0.1255	0.4320
Dataset	Rent listings of residential properties (REIDIN)					
Metro effect	0.049*** (0.013)	0.007 (0.035)	0.084*** (0.014)	0.083** (0.034)	0.209*** (0.020)	0.119*** (0.027)
R^2 *within*	0.1040	0.2349	0.1042	0.2291	0.1051	0.2362
Dataset	Sale transactions of retail properties (RERA)					
Metro effect	0.398*** (0.054)	-0.207 (0.148)	0.426*** (0.037)	0.132 (0.106)	0.409*** (0.036)	0.441** (0.194)
R^2 *within*	0.1253	0.1355	0.1467	0.1340	0.1424	0.1476

Legend: RCS: repeated cross-sectional data. PP: Pseudo panel data

*, **, *** indicate significance at 10%, 5% and 1% respectively. Values in parentheses are standard errors

Note: Although other variables as listed in Table 6-4 (e.g. property characteristics, neighbourhood characteristics, distance to amenities) are included and estimated in the model, the results above only present the impact of the Dubai Metro, since this is the main purpose of this research. The results for the contextual variables are available upon request.

Model type	First version of the HP models					
Equation number	**19**	**20**	**19**	**20**	**19**	**20**
Catchment zone	**0.5 km**		**1 km**		**1.5 km**	
Data structure	**RCS**	**PP**	**RCS**	**PP**	**RCS**	**PP**
Dataset	Sale transactions of residential properties (RERA)					
Metro effect	-0.096*** (0.013)	-0.066 (0.059)	0.016* (0.010)	-0.022 (0.041)	0.009 (0.005)	-0.025 (0.021)
R^2 *within*	0.1935	0.1260	0.1924	0.1256	0.1924	0.1261
Dataset	Sale listings of residential properties (REIDIN)					
Metro effect	-0.012** (0.006)	-0.015 (0.021)	0.019*** (0.004)	0.006** (0.002)	0.026*** (0.004)	0.010 (0.008)
R^2 *within*	0.1251	0.4298	0.1252	0.4296	0.1251	0.4302
Dataset	Rent listings of residential properties (REIDIN)					
Metro effect	0.053*** (0.013)	0.009 (0.034)	0.029*** (0.006)	0.048* (0.025)	0.015*** (0.003)	0.004*** (0.004)
R^2 *within*	0.1040	0.2347	0.1041	0.2396	0.1041	0.2346
Dataset	Sale transactions of retail properties (RERA)					
Metro effect	0.160*** (0.061)	-0.019 (0.132)	0.303*** (0.041)	0.131* (0.068)	0.102*** (0.022)	0.093* (0.052)
R^2 *within*	0.1552	0.1777	0.1674	0.1910	0.1588	0.1893

Legend: RCS: repeated cross-sectional data. PP: Pseudo panel data

*, **, *** indicate significance at 10%, 5% and 1% respectively. Values in parentheses are standard errors

Note: Although other variables as listed in Table 6-4 (e.g. property characteristics, neighbourhood characteristics, distance to amenities) are included and estimated in the model, the results above only present the impact of the Dubai Metro, since this is the main purpose of this research. The results for the contextual variables are available upon request.

GC type	GC of public transport		Weighted average GC	
Equation number	**24**	**25**	**26**	**27**
Data structure	**RCS**	**PP**	**RCS**	**PP**
Dataset	Sale transactions of residential properties (RERA)			
GC (0.5 km)	0.002*** (0.0005)	0.002 (0.003)	-0.009 (0.007)	-0.015** (0.031)
GC (>0.5-1 km)	-0.0003 (0.0006)	0.002 (0.003)	0.022*** (0.005)	0.004 (0.023)
GC (>1-1.5 km)	-0.0003** (0.0002)	-0.001 (0.003)	0.0007 (0.008)	-0.016 (0.018)
GC (>1.5 km)	0.001*(0.0007)	-0.0006 (0.001)	0.043*** (0.006)	0.014 (0.014)
R^2 *within*	0.1930	0.1262	0.1942	0.1271
Dataset	Sale listings of residential properties (REIDIN)			
GC (0.5 km)	-0.002*** (0.001)	-0.001 (0.002)	-0.058*** (0.005)	-0.047*** (0.001)
GC (>0.5-1 km)	-0.004*** (0.0004)	-0.003 (0.002)	-0.078*** (0.003)	-0.049*** (0.012)
GC (>1-1.5 km)	-0.005*** (0.0001)	-0.009 ** (0.003)	-0.086*** (0.010)	-0.036 (0.038)
GC (>1.5 km)	-0.002*** (0.0001)	-0.001 (0.001)	-0.056*** (0.002)	-0.057*** (0.010)
R^2 *within*	0.1262	0.4302	0.1295	0.4447
Dataset	Rent listings of residential properties (REIDIN)			
GC (0.5 km)	-0.012*** (0.001)	-0.007* (0.004)	0.065*** (0.024)	0.068 (0.070)
GC (>0.5-1 km)	-0.010*** (0.001)	-0.009** (0.004)	-0.049*** (0.008)	0.013 (0.063)
GC (>1-1.5 km)	-0.010*** (0.002)	-0.012** (0.005)	-0.031*** (0.037)	0.086 (0.103)
GC (>1.5 km)	-0.011*** (0.001)	-0.0007 (0.005)	0.039*** (0.004)	-0.008 (0.029)
R^2 *within*	0.1056	0.2459	0.1056	0.2373
Dataset	Sale transactions of retail properties (RERA)			
GC (0.5 km)	-0.0007 (0.003)	0.013***(0.003)	0.164*** (0.038)	0.051 (0.066)
GC (>0.5-1 km)	-0.009** (0.003)	-0.0003 (0.005)	0.276*** (0.049)	0.126 (0.089)
GC (>1-1.5 km)	-0.00001 (0.013)	-0.006 (0.006)	0.083** (0.038)	0.060 (0.076)
GC (>1.5 km)	0.050** (0.019)	-0.008 (0.017)	-0.003 (0.186)	-0.302 (0.369)
R^2 *within*	0.1629	0.1825	0.1654	0.1994

Legend: RCS: repeated cross-sectional data. PP: Pseudo panel data

*, **, *** indicate significance at 10%, 5% and 1% respectively. Values in parentheses are standard errors

Note: Although other variables as listed in Table 6-4 (e.g. property characteristics, neighbourhood characteristics, distance to amenities) are included and estimated in the model, the results above only present the impact of the Dubai Metro, since this is the main purpose of this research. The results for the contextual variables are available upon request.

Appendix E – The results on an annualized basis

Table 11-9: The results on an annualized basis for the effect of the Dubai Metro on property values

Catchment zone		0.5 km				1 km				1.5 km			
Data structure		RCS		PP		RCS		PP		RCS		PP	
Model type		DID Basic	DID Full	DID Basic	DID Full	DID Basic	DID Full	DID Basic	DID Full	DID Basic	DID Full	DID Basic	DID Full
Sale transactions of residential properties	Model result	-0.057*** (0.014)	-0.090*** (0.014)	-0.036 (0.051)	-0.009558	0.108*** (0.007)	0.078*** (0.008)	0.055 (0.034)	0.043 (0.039)	0.101*** (0.010)	0.018 (0.013)	0.084*** (0.030)	0.033 (0.038)
	Result on an annualized basis	-0.023	-0.036	-0.014	-0.071	0.043	0.031	0.022	0.017	0.040	0.007	0.074	0.013
sale listings of residential properties	Model result	-0.013** (0.006)	-0.014** (0.006)	0.026 (0.021)	0.009 (0.021)	0.031*** (0.004)	0.026*** (0.005)	0.069*** (0.017)	0.017 (0.017)	0.046*** (0.005)	0.026*** (0.004)	0.160*** (0.019)	0.089*** (0.019)
	Result on an annualized basis	0.009	-0.009	0.017	0.006	0.021	0.017	0.046	0.011	0.031	0.017	0.107	0.059
Rent transactions of residential properties	Model result	0.051*** (0.013)	0.051*** (0.013)	0.081** (0.032)	0.037 (0.027)	0.026*** (0.006)	0.010* (0.006)	0.249*** (0.026)	0.087*** (0.023)	0.025*** (0.006)	0.008 (0.006)	0.351*** (0.027)	0.116*** (0.025)
	Result on an annualized basis	0.034	0.034	0.054	0.025	0.017	0.007	0.166	0.058	0.017	0.005	0.234	0.077
Sale transactions of commercial properties	Model result	0.171	0.394	0.115	0.166	0.406	0.42	0.273	0.303	0.417	0.405	0.444	0.458
	Result on an annualized basis	0.068	0.158	0.046	0.066	0.162	0.168	0.109	0.121	0.167	0.162	0.178	0.183

Catchment zone		0.5 km		1 km		1.5 km	
Data structure		RCS	PP	RCS	PP	RCS	PP
Model type		First version of HP model	First version of HP model	First version of HP model	First version of HP model	First version of HP model	First version of HP model
Sale transactions of residential properties	Model result	-0.086*** (0.013)	-0.031 (0.051)	0.019* (0.010)	0.004 (0.034)	0.012** (0.005)	0.002 (0.015)
	Result on an annualized basis	-0.017	-0.006	0.004	0.001	0.002	0.00040
sale listings of residential properties	Model result	-0.014** (0.006)	-0.011 (0.020)	0.015*** (0.004)	0.005 (0.012)	0.004*** (0.002)	0.019** (0.007)
	Result on an annualized basis	-0.005	-0.004	0.005	0.002	0.001	0.006
Rent transactions of residential properties	Model result	0.041*** (0.012)	-0.009 (0.028)	0.016** (0.007)	0.027 (0.018)	0.010*** (0.004)	0.030*** (0.011)
	Result on an annualized basis	0.014	-0.003	0.005	0.009	0.003	0.010
Sale transactions of commercial properties	Model result	0.356*** (0.052)	0.234** (0.108)	0.336*** (0.029)	0.245*** (0.058)	0.138*** (0.014)	0.175*** (0.037)
	Result on an annualized basis	0.071	0.047	0.067	0.049	0.028	0.035

Dataset	Sale transactions of residential properties	Sale listings of residential properties	Rent listings of residential properties	Sale transactions of retail properties
Data structure	**RCS**	**RCS**	**RCS**	**RCS**
Model type	**Second version of HP model**	**Second version of HP model**	**Second version of HP model**	**Second version of HP model**
Model result				
Less or equal to 300 m	**-0.016 (0.025)**	**-0.004 (0.009)**	**0.086*** (0.022)**	**0.203*** (0.061)**
301 - 500 m	**-0.045** (0.020)**	**0.014** (0.007)**	**0.025* (0.015)**	**0.381*** (0.073)**
501 - 700 m	**0.015 (0.020)**	**0.044*** (0.006)**	**-0.013 (0.009)**	**0.395*** (0.071)**
701 - 900 m	**0.097*** (0.019)**	**0.018*** (0.006)**	**0.095*** (0.019)**	**0.745*** (0.075)**
901 - 1100 m	**0.075*** (0.018)**	**0.069*** (0.008)**	**0.034* (0.019)**	**0.378*** (0.091)**
1101 - 1300 m	**-0.016 (0.016)**	**0.054*** (0.011)**	**0.019 (0.022)**	**0.284*** (0.083)**
1301 - 1500 m	**-0.010 (0.019)**	**0.018 (0.022)**	**0.059 (0.081)**	**0.273** (0.122)**
Result on an annualized basis				
Less or equal to 300 m	**-0.003**	**-0.001**	**0.029**	**0.041**
301 - 500 m	**-0.009**	**0.005**	**0.008**	**0.076**
501 - 700 m	**0.003**	**0.015**	**-0.004**	**0.079**
701 - 900 m	**0.019**	**0.006**	**0.032**	**0.149**
901 - 1100 m	**0.015**	**0.023**	**0.011**	**0.076**
1101 - 1300 m	**-0.003**	**0.018**	**0.006**	**0.057**
1301 - 1500 m	**-0.002**	**0.006**	**0.020**	**0.055**

GC type		GC of public transport		Weighted average GC		GC of public transport		Weighted average GC	
Data structure		**RCS**	**PP**	**RCS**	**PP**	**RCS**	**PP**	**RCS**	**PP**
		Model result				Result on an annualized basis			
Sale transactions of residential properties	**GC (0.5 km)**	**0.002*** (0.0005)**	**0.0001 (0.001)**	**-0.002 (0.003)**	**-0.038*** (0.009)**	0.0004	0.00002	-0.0004	-0.0076
	GC (>0.5-1 km)	**-0.0002 (0.0005)**	**0.002 (0.001)**	**-0.0007 (0.002)**	**-0.032*** (0.009)**	-0.00004	0.002/5	-0.00014	-0.0064
	GC (>1-1.5 km)	**-0.0003** (0.0002)**	**0.002 (0.001)**	**-0.003*** (0.0006)**	**-0.035*** (0.010)**	-0.00006	0.0004	-0.0006	-0.007
	GC (>1.5 km)	**-0.004*** (0.0004)**	**-0.003*** (0.0009)**	**-0.020*** (0.003)**	**-0.053*** (0.010)**	-0.0008	-0.0006	-0.004	-0.0106
sale listings of residential properties	**GC (0.5 km)**	**-0.001*** (0.0005)**	**-0.0002 (0.0007)**	**-0.024*** (0.002)**	**-0.0007 (0.002)**	-0.0003	-0.00007	-0.008	-0.0002
	GC (>0.5-1 km)	**-0.002*** (0.0004)**	**-0.000003 (0.0007)**	**-0.033*** (0.002)**	**-0.0010 (0.002)**	-0.0007	-0.000001	-0.011	-0.0003
	GC (>1-1.5 km)	**0.0007 (0.0006)**	**0.004*** (0.001)**	**-0.021*** (0.003)**	**0.006* (0.003)**	0.0002	0.001	-0.007	0.002
	GC (>1.5 km)	**-0.003*** (0.0002)**	**-0.003*** (0.0005)**	**-0.037*** (0.002)**	**-0.027*** (0.003)**	-0.001	-0.001	-0.012	-0.009
Rent transactions of residential properties	**GC (0.5 km)**	**-0.006*** (0.0008)**	**-0.003*** (0.0008)**	**-0.007*** (0.002)**	**-0.006*** (0.002)**	-0.002	-0.001	-0.002	-0.002
	GC (>0.5-1 km)	**-0.004*** (0.0006)**	**-0.002*** (0.0008)**	**-0.009*** (0.002)**	**-0.005** (0.002)**	-0.001	-0.0007	-0.003	-0.002
	GC (>1-1.5 km)	**-0.006*** (0.001)**	**-0.003*** (0.001)**	**-0.008*** (0.002)**	**-0.006** (0.002)**	-0.002	-0.001	-0.0026667	-0.002
	GC (>1.5 km)	**-0.006*** (0.0007)**	**-0.0008 (0.001)**	**0.014*** (0.002)**	**-0.003 (0.003)**	-0.002	-0.0003	0.005	-0.001

	GC (0.5 km)	**-0.012*** (0.002)**	**-0.009** (0.004)**	**0.108*** (0.016)**	**0.017 (0.020)**	-0.0024	-0.0018	0.0216	0.0034
	GC (>0.5-1 km)	**-0.017*** (0.002)**	**-0.010** (0.004)**	**0.113*** (0.018)**	**0.017 (0.022)**	-0.0034	-0.002	0.0226	0.0034
	GC (>1-1.5 km)	**-0.018*** (0.003)**	**-0.015*** (0.004)**	**0.103*** (0.019)**	**0.003 (0.024)**	-0.0036	-0.003	0.0206	0.0006
Sale transactions of commercial properties	**GC (>1.5 km)**	**-0.025*** (0.004)**	**-0.024*** (0.005)**	**0.078*** (0.022)**	**-0.029 (0.027)**	-0.005	-0.0048	0.0156	-0.0058