University of Glasgow

School of Social and Political Sciences

Urban Studies

The Analysis of Six Selected Factors Affecting Ridership in 159 Metro Systems Worldwide

Presented in partial fulfilment of the requirements for degree of

M.Sc. in Urban Transport

August 2017

Word count: 14,157

Abstract

Nowadays, cities all over the world, in both developed and developing countries, have striven to fulfil the sustainable standards of urban transport by improving public transport, encouraging non-motorized modes, limiting the use of private cars. Featuring a considerably higher combination of passenger volume and speed per unit of railway space than any other mode, metros become a popular means of mass public transport around the world. In order to improve the operation efficiency of metro systems, corresponding researches should be expanded to better understand the influential determinants of metro ridership per capital. The objective of this study is to examine factors influencing metro ridership per capital, with the goal of supplementing planning and policy decisions. Six main factors including system location with specific continent, country's economic activity level, city population, system accessibility, age of system and the presence of BRT or LR systems are introduced to explain the variance in annual ridership per capital of 159 metro systems worldwide for the year 2017. The findings could provide support for international institutions such as the World Bank, and other investment banks such as Inter-American Development Bank, Asian Development Bank etc. in their decision-making and target- testing process.

Table of Contents

List of Figures

List of Tables

Acknowledgments

The dissertation period is one of considerable time and effort. I would be remiss if I failed to offer my sincere thanks to those who assisted me in this long and arduous task. Firstly, I would like to thank professor Piyushimita Thakuriah for her endless patience and guidance as my supervisor. It would not have been possible to complete this study without her. Her extraordinary support guided me throughout this year and I will be forever grateful to her for this. I am very appreciative of all the staff from the Urban Big Data Centre for their thoughtfulness in both the curriculum and topics selected to keep me constantly engaged. I am also indebted to the support offered by the World Metro Database, which give me access to the dataset that was the foundation of this study. My thanks to my classmates for bearing with me during these past few months, who were always there for me to brainstorm concepts and hash out ideas. Finally, I would like to express my profound gratitude to my family for providing with continuous encouragement through this process. I could not have gotten to where I am today without my parents, they are the calm in the storm and the light that I strive towards when I lost.

1. Introduction

Urban transport is an extremely vital component of urban sustainability as it has significant economic, social and environmental impacts (Puche and Lefèvre, 1996; Pojani and Stead, 2015; Gwilliam, 2002; Haghshenas and Vaziri, 2012; Gilbert et al.,2003; Goldman and Gorham, 2006). Nowadays, cities all over the world, in both developed and developing countries, have striven to fulfil the sustainable standards of urban transport by improving public transport, encouraging non-motorized modes, limiting the use of private cars (De Vos and Witlox, 2013; Meyer et al.,1996; Ardila, 2007; Cornette, 2004; Richardson, 2005). Specific solutions including pull measures and push measures, for instance, encouraging residents to use public transport, improving the infrastructure for cyclists and pedestrians, subsidizing renewable fuel, increasing the price for parking, raising the tax on fossil fuel and charging the congestion fee on private cars in the inner city, are promoted to solve a series of traffic problems caused by automobile dominance (Eriksson et al., 2008; Topp and Pharoah, 1994; Dablanc, 2007). Urbanization also shows an inevitable trend given that global cities are growing in both scale and quantity (Kasarda and Crenshaw, 1991; Barter, 1999; Cohen, 2006; Anderson et al., 2005). An estimated 54.5 percent of world's population resided in urban settlements in 2016, and urban areas are projected to accommodate 60 percent of inhabitants globally by 2030 (United Nations, 2016). Large population concentrated in urban areas would intensify the levels of tension of resources conflicts between private transport and public transport and incur more severe traffic congestion, road accidents and environmental pollution (Cornette, 2004; Gwilliam, 2002). Measures with regard to developing public transport are high on the agenda in many global cities, particularly in urban areas with relatively higher population density and intensive land use, often including buses, trams or mass rapid transport modes. (Derrible and Kennedy, 2010; Pucher, 1988; Pojani and Stead, 2015; Wright and Fjellstrom, 2003; Garrett and Taylor, 1999.)

Accompanying with urbanization, transport corridors are spontaneously and gradually evolving in many megacities and large cities worldwide , metros as the representative of mass rapid transport modes can sustain a correspondingly high level of corridor activity (Mohammadi et al., 2018; Freilich and Chinn, 1986; Loo et al., 2010; Jabareen, 2006) This mode can greatly fulfil the ever-increasing travelling demand, effectively alleviate land pressure in city centres via transporting commuters from dominant areas where centralize considerable commercial activities and employments opportunities to peripheral areas (Levinson, 2000). The last 15 years have witnessed a considerable expansion of metro systems that a total of 53 new metro systems were built and put into operation since the millennium (UITP, 2015). The latest statistics of global metro figures displays that up to 2014 alone, there are 157 cities around the world with a metro system in operation, 513 km of new metro infrastructure and 355 new metro stations were put into service (see figure 1). Interestingly noted that nearly two thirds of these networks are located in Asia and Europe, with the number of 54 and 46 respectively, 18 systems in Latin America, 16 in both Eurasia and North America, and 7 in the Middle East and North Africa region. Figure 1 shows a map of global countries with metro systems.

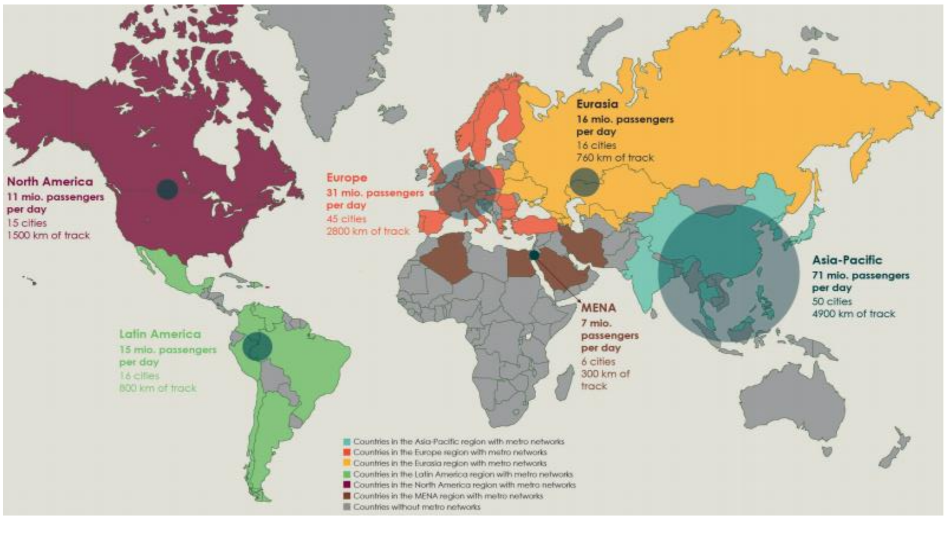


Figure 1.1 Map of countries having metro networks according to world region

Source: UITP transport

Compared with Bus Rapid Transport and light rails, metros have the prevailing performance featuring a considerably higher combination of passenger volume and speed per unit of railway space than any other mode (Ben-Akiva and Morikawa, 2002; Demery, 1994; Mohammadi et al., 2008; Levinson, 2000). Implement of metro systems usually play a crucial role in the movement of millions of passengers, the carrying capacity of metros is generally designed around 30,000 to 40,000 passengers per hour in the peak direction requiring a city with a population of 2-3 million having at least one transport corridor (Mitric, 1997; Duncan, 2010). Another competitive advantage over other public transport is high speed as the result of an electrified third rail while is fatal on contact, hence fully segregated rights-of-way are always request. The fully segregated systems extremely avoid road congestion and possess outstanding quality of service in terms of punctuality, comfort and safety (Pojani and Stead, 2015). Additionally, travelling by metro can effectively reduce tailpipe emissions when they attract potential motorists, particularly single-occupant drivers (Mohammadi et al., 2008).

The flip side of these advantages is high capital and operation cost, which makes them less economically viable in medium-sized developing cities than in megacities. The capital cost of metro systems varies between cities, between systems, and between metro lines within the same system due to ground conditions, the ratio of underground or above-ground construction, type of rolling stock and so on, which is approximately \$50–\$150 million per kilometre with cost over-runs being the rule (Sinfield and Einstein, 1998). Besides high upfront costs, the operating cost is also substantially high, which usually require operating subsidies, otherwise the price of the tickets would be prohibitive for local residents even in developed cities (Guerra and Cervero, 2011; Crotte et al., 2011). Additionally, planning and guidelines focus more on benefits valuation than cost estimates (Guerra, 2011; Flyvbjerg et al., 2013; Levinson, 2000; Israel and Cohen-Blankshtain, 2010). Considering the evolution of spatial distribution of population and economic activities usually fail to accurately predict, metro systems are often designed within a time-limited horizon, while adjustments and correlations are hard to make on the fixed railways and stations once construction completed (Roth et al., 2012). On another hand, Guerra and Cervero (2011) promoted that high costs and low ridership is the bane of overall deficit of metro systems globally. Expensive transport projects need high ridership levels (Garrett and Taylor, 1999). A metro system with higher ridership means that more passengers can be served with the same resources, meanwhile more fares collected and lower vehicle miles travelled. If metros systems could not achieve reasonable ridership in order to guarantee operation efficiency, they will also fail to produce substantial environmental or social benefits. Hence, understanding the relative factors on metro ridership is central to public policy debates over metro project investments.

The traditional literature on influencing factors of transport ridership is abundant. Studies of these influential determinants can be spilt into two categories, researches focusing on travelling attitudes and perceptions and researches examining travelling patterns, built environment and system characteristics associated with ridership (Taylor and Fink, 2003; Kohn, 1999; Kain and Liu, 1999; Gomez-Ibanez, 1996; Taylor et al., 2009; Taylor et al., 2013). Corresponding influential factors can be broadly divided into external factors and internal factors. External factors, such as the level of funding, employment levels, income levels, auto ownership and urban form, are largely exogenous to transport systems and their managers. Internal factors are those over which transport managers can exercise some control, usually including fare innovation and changes, marketing and information, new planning approaches and partnerships, service quality and quantity. However, past literatures on factors affecting metro transport is uneven, in some cases poorly conceived (Ben-Akiva and Morikawa, 2002; Vuchic and Musso, 1991; Derrible and Kennedy, 2010; Derrible and Kennedy, 2009; Zhang et al., 2011; Angeloudis and Fisk, 2006; Bhandari et al., 2009; Barberillo and Saldana, 2011). Most recent literatures about factors influencing metro ridership mainly focus on station-level (Chan and Miranda-Moreno, 2013; Lee et al., 2008; Lin and Shin, 2008; Zhao et al., 2013; Jun et al., 2015; Sohn and Shim, 2010; Lee et al., 2013). The influence of these factors on metro ridership can be taken into consideration for direct strategies and indirect strategies of public transport policies in order to increase operation efficiency and effectiveness, encourage public transport use.

Attention to influential determinants on ridership from the perspective of entire metro systems is the area of concern for our research. A total of 159 global metro systems were selected as study areas. The dataset was carefully constructed using a number of data sources, including data published by the operating agencies. A wide arena of metro systems was observed over six major world areas including all network sizes (from 1 to 25 lines), thereby considering different cultures and specificities as well as the patterns of developments of network systems. Moreover, we restricted our analysis to ridership per capital given that transport ridership per capital can signify individual preference of travelling choice which is useful at the strategic planning phase and setting long-term goals of operation and management. Ridership data available for the latest year 2017 also adds the advantage of unique data for this study.

A multiple regression model has been set up to examine how six selected factors explain the variance in annual ridership per capital of 159 global metro systems for the year 2017. The Six main factors are relating to metro system location with specific continent, country's economic activity level, city population, the presence of BRT or LR systems, system accessibility, age of system. Continent is introduced as a group of explanatory variables in five categories to describe the specific continent where metro systems located. Income is introduced to measure the average income levels from country level in four categories to estimate economic influence on personal metro ridership. City is introduced as a group of variables in five categories based on the amount of city population. The presence of BRT or LR systems in cities with metro systems is termed as MRT. The influence of system accessibility on metro ridership per capital is measured by a group of variables termed as Form in three categories. The operating years of 149 metro systems are collected as Years for analyzing the effect of age of systems.

More specifically, the study attempts to answer these question:

1) How does system location with specific continent affect annual metro ridership per capital when controlling with Income, City, MRT, Form and Years?
2) How does country's economic activity level affect annual metro ridership per capital when controlling with Continent, City, MRT, Form and Years?
3) How does city population affect annual metro ridership per capital when controlling with Continent, Income, MRT, Form and Years?
4) Does the presence of BRT or LR systems have any influence on annual metro ridership per person when controlling with Continent, Income, City, MRT, Form and Years?
5) How does system accessibility affect annual metro ridership per capital when controlling with Continent, Income, City, MRT and Years?
6) How does age of system affect annual metro ridership per capital when controlling with Continent, Income, City, MRT and Form?

The objective of this study is to examine factors that influence ridership of metro systems, with the goal of supplementing planning and policy decisions. This study offers a means to compare metro systems worldwide in order to help local planners and agencies in their decision-making and target-testing process, as well as to inform international institutions such as the World Bank, and other investment banks such as Inter-American Development Bank, Asian Development Bank etc. The overall purpose of the research is to understand how these factors are potentially related to metro ridership and the corresponding consequent effect on metro systems design and planning, thereby achieving a comprehensive and comparative understanding of metro systems.

The structure of this study is as follow. Section 2 will consist of a literature review of factors influencing metro ridership. In section 3, an overview of the methodology will be presented, followed by a description of study area, data collection approach, variables used and the multiple linear regression model that used to answer those research questions. Section 4 will then examine the results from the regression models. Section 5 will consist of a discussion of the results along with corresponding policy implications. Finally, section 6 will offer conclusions and recommendations for future research.

2. Literature review

This section will first review influential determinants of metro ridership, and then influential factors on metro ridership per capital are generally divided into internal and external factors. Afterwards, literatures on external factors will be reviewed to inspect the influence of socio-economic factors, spatial factors and public finance factors on metro ridership per capital. Finally, the influence of internal factors from perspectives of service quantity and service quality will be reviewed and discussed in detail.

2.1 Main Factors on metro ridership

The traditional literature on influencing factors of transport ridership is abundant (Taylor et al., 2009; Taylor and Fink, 2003; Pickrell, 1989; Kain and Liu, 1999; Cervero, 1990; Kuby et al., 2004; Paulley, et al., 2006; Litman, 2004; Watkins et al.; Fujii et al., 2011; Rissel et al., 2012; Hine and Scott, 2000; Gronau and Kagermeier, 2007; Zhan et al., 2007; Cervero, 1994; Ben-Akiva and Morikawa, 2002; Hess and Almeida, 2007; Mackett and Edwards, 1998). Studies of the influential determinants of transport ridership can be spilt into two categories: first, studies that focus on travelling attitudes and perceptions, in which travelers or operators are viewed as the unit of analysis; second, studies that examine travelling patterns, built environment and system characteristics associated with transport ridership, including both disaggregated studies based on individual transport mode choice and aggregate studies where transport systems are viewed as the unit of analysis and independent variables are usually in metropolitan level. Specific to public transport ridership, a literature review of ridership enhancement on public transport projects by European Commission (1996) outlined a list of factors including facilities, marketing, use restrictions (e.g., road pricing, parking costs, access restrictions), technologies for providing services, changes in the fare levels, service quality and quantity, taxes on car ownership and other policies from land use planning.

A host of influencing factors are examined by descriptive analysis and causal analyses. Descriptive analyses usually use survey data to assess perceptions of the factors affecting ridership. Causal analyses posit and test hypotheses about the factors influencing transport ridership. Many of these studies used multivariate regression analysis to identify the factors most strongly related to changes in transport ridership. An import methodological issues is the generally high levels of collinearity among the various spatial variables, and among the spatial variables and many socio-economic variables. Another unfortunate commonality between many of the previous studies, small sample sizes, raises questions about both the generalizability and statistical significance of findings. Furthermore, the broad conceptual factors hypothesized to influence ridership and the variables operationalized in these models vary widely (Holmgren, 2007; Taylor and Fink, 2003).

More recently, a large growing body of researches on metro systems appears (Nieuwenhuijsen et al., 2007; Derrible and Kennedy, 2010; Boccaletti, 2006; Yang et al.,2013; Siemiatycki, 2006; González-Gil et al., 2013; Li and Lo, 2014; Grava, 2003; Pagliara and Papa, 2011). A fundamental study on approaching metros as potential development projects was conducted by Mitric (1997), in which metro systems are seen as effective responses to problems linked to increasing motorization, poor road infrastructure, spatial patterns with large passenger volume, better quality of service in terms of punctuality, comfort and safety than any other mode. Roth et al. (2012) looked at the temporal evolution of metro systems in an exploratory manner of structures, pointed that all the world's largest metro networks converge to a shape that shares similar generic features despite their geographical and economic differences, which is beneficial to summarize the characteristic of metro systems from a global view.

However, the specific studies of influencing factors on metro ridership are surprisingly limited and uneven in terms of data, methods or findings, although it is clearly an important area of public transport policy research (Vuchic and Musso, 1991; Angeloudis and Fisk, 2006; Bhandari et al., 2009; Lin and Shin, 2008. Loo et al., 2010; Dill et al,, 2013; Allport, 1981). Substantial studies of factors on metro ridership mainly focus on station level (Lee et al., 2008; Zhang et al., 2011; Barberillo and Saldana, 2011; Zhao et al., 2013; Jun et al., 2015; Sohn and Shim, 2010; Lee and Hong, 2013.).

An investigation into the factors affecting metro demand at a station level in the Seoul metropolitan area was carried by Sohn and Shim (2010). A regression analysis was conducted with weekly average of station boarding as the dependent variable. The 24 independent variables were chosen and categorized into three groups: built environment, external connectivity, and intermodal connection. Seven variables proved to be significantly associated with station boarding passenger: employment rate, commercial floor area, office floor area, net population density, number of transfers, number of feeder bus lines and a dummy variable indicating transfer stations. Variables of land use and walkability proved to be indirectly related to station boarding passenger through employment rate. Several separate links represented a cyclic relationship between external connectivity and intermodal connection, the number of feeder lines was found to have a reciprocal relationship with station boarding. Jun et al. (2015) evaluated built environments surrounding Seoul's metro stations in terms of transit-oriented development principles and to examine their influence on metro ridership. Population and employment densities, land use mix diversity, and intermodal connectivity indicated by the number of stations were positively related to metro ridership, which is consistent with the results of the existing studies. The major contribution of this study is to find that employment density and stations connectivity affects metro ridership over wider spatial ranges than do population density and level of mixed land use.

A cross-section analyses of 46 metro stations in Taipei City, Taiwan, China for 2004 performed by Lin and Shin (2008) showed that daily ridership was positively affected by the floor-space area of the station areas and insignificantly affected by mixed land use. Ridership dispersion in time was positively influenced by sidewalk length, negatively affected by retail and service floor-space area, and insignificantly influenced by density. Zhao et al. (2013) investigated the impacts of land use, external connectivity, intermodal connection, and station context on ridership of 55 metro stations in China, which is considered as one of the most ambitious metro transit expansions in the world, by Geographic Information System and multiple regression models. Six variables were found to be significantly associated with metro station ridership at the 0.05 level: population, business/office floor area, CBD dummy variable, number of education buildings, entertainment venues and shop centers. Five variables were proved to be related to station ridership at the 0.01 significance level: employment, road length, feeder bus lines, bicycle park-and-ride spaces, and transfer dummy variable. Lee et al. (2013) indicated that metro riders in the station areas located in the CBD area and the fringe areas or the periphery are influenced mostly by density, whereas metro riders in the station areas located in the sub-central areas or the inner and outer suburbs are affected mainly by diversity. The results provide policy implications for transit-oriented development strategies to increase metro ridership in metropolitan areas where a number of sub centers exist.

Given that lots of recent studies focusing a combination of internal and external factors by regression analysis, it is interesting to note that external factors, for instance, income, parking policies, development, employment, fuel prices, car ownership, and density levels, are found to have greater effects on ridership than internal factors. Of the internal factors, service quality is often found to be more important than low fares (Gomez-Ibanez, 2013). However, it is important to note that there is no hard line separating internal from external factors, for example, increased population growth may change the demand for transit services which in turn may change the levels of service provision (Taylor et al., 2009; Taylor and Fink, 2003).

To summarize, factors influencing metro ridership are broadly divided into external factors and internal factors. External factors as proxies for large numbers of factors thought to affect transport demand are largely exogenous to systems and their managers, which can be further split into three categories as socio-economic factors, spatial factors and public finance factors, including employment levels, income levels, auto ownership, the price of gasoline and parking costs, parking strategies, urban form (residential densities and employment densities) and the level of funding. Internal factors are those over which transport managers could exercise some control, such as service improvements and adjustments, fare innovation and changes, marketing and information, new planning approaches and partnerships, service quality and coordination.

2.2 Selected factors on metro ridership

2.2.1 External factors on metro ridership

2.2.1.1 Socio-economic factors and spatial factors

A combination of variables based on socio-economic characteristics, demographic characteristics, land use, inter-modal competition from different dimensions was important in accounting for the variability of metro ridership. Dill et al. (2013) examined the combined influence of urban form and service levels on metro ridership at the station level. Categories of independent variables tested include: socio-demographics; land use (population, employment, land use type etc.); service levels (transfer stops, hours of service, headways, bus and light rail etc.). The final results of regression model indicate socio-demographic characteristics have a larger effect on metro ridership in the large urban area than small urban areas, while land use characteristics have much smaller effect in large urban area than small urban area. Land use characteristics around metro stations do have significant effects on ridership, though these effects are much smaller than the effects of service level.

Since policy makers have some direct control over the development of transport systems and land use, while less control over many of socio-economic factors (Taylor and Fink, 2003), there are several recent researches on the relationship between urban form and metro ridership. Durning and Townsend (2015) produced a ridership model in context of Canadian rapid transport. by using station boarding as the dependent variable and 44 socio-economic, built-environment and system attributes as potential explanatory variables, yielded one model with an adjusted R^2 value of 0.8033. The results are similar to those of models constructed in the United States with respect to densities, land uses and station amenities, and socioeconomic variables do not appear to be significant. The case studies of metro systems in New York City and Hong Kong by Loo et al. (2010) show that place-specific factors are important in influencing metro ridership and car ownership, which are both significant and positively associated with average weekday metro ridership per capital. Guerra and Cervero (2011) found that population and job densities are positively correlated with both metro ridership and capital costs when controlling for neighborhood, regional and transport service attributes. Increasing the number of jobs and residents around existing metro stations and limiting new capital investments to transport-supportive areas are essential for increasing metro ridership while containing costs.

Pojani and Stead (2015) pointed that it is debatable in medium-sized developing cities whether rail-based or road-based public transport should be emphasized, given that they are more flexible in terms of urban expansion, adoption of green travel modes and environmental protection. They also pointed that light rail transport is generally more appealing to middle class passengers, investment of this mode is seen as a signal of a more permanent commitment of government to public. Chalermpong (2007) prompted that either rail- or road-based public transport might not be economically viable in urban areas with a small but dispersed population. Ben-Akiva and Morikawa (2002) pointed that riders' preferences for rail over bus is null when service levels such as travel time or cost are equal according to a few quantitative analyses, and the preferences only arise when one mode offers a higher quality service such as fewer transfers or higher frequency. Kamruzzaman (2013) pointed that evolving technologies (e.g., electric buses) have minimized the differences between bus and rail in terms emissions, capacity and comfort.

A global transport review conducted by the World Bank (Gwilliam, 2002), Cities on the move, provided a series of strategies for national and city governments to address urban transport problems, particularly the role and limitations of Mass Rapid Transport (MRT) in chapter 8. AchARyA and MoRichi (2007) argued that MRT can play a significant role in improving overall condition of urban transport. They also pointed that the impact of MRT would be insignificant if the investment is committed too late. Wright and

Fjellstrom (2003) promoted that although there is no single MRT solution fitting all cities, for all but the major corridors of relatively wealthy and dense developing cities, the best option to develop an MRT system will often be a form of Bus Rapid Transport. They also pointed that metros and light rail transit are still relatively uncommon in low income developing cities. Ardila (2007) found that developing cities often lack the institutional capacity to simultaneously develop multiple transport system. In practice, especially in developing countries, once a particular MRT system develop, resources tend to be devoted to that system, while other transport mode are neglected, although theoretically cities should follow a balances approach by using complementary MRT systems appropriate to local circumstance. Daganzo (2010) found that Bus Rapid Transport (BRT) effectively competes with the automobile, and also outperform metro if a city has enough suitable streets on which to run BRT, even the city is large and the demand high. They also pointed that since power shortages are common in many developing cities, metro systems which rely on grid electricity may not always be a feasible or desirable option. Gomez-Ibanez (1985) prompted that it is crucial for medium-sized cities to construct bus rapid transport rather than light rails or metros. Additionally, Vuchic and Musso (1991) pointed that corrections are extremely difficult to make on built metro lines and networks, which are fixed, relatively permanent and have extensive infrastructure, compared with bus networks since bus routes are very dependent on local conditions and can be easily modified

2.2.1.2 Public finance factors

Gomez-Ibanez (1996) found that large transport subsidies have led to negligible ridership growth and no increase in the proportion of daily travelling trips making by mass public transport modes in the United States. Empirical evidence in Western Europe and North America had suggested that the decentralization of transport finance and decision making may enhance the effectiveness of subsidies. Pucher and Kurth (1995) conducted a research on how to deal most effectively with limited subsidy funds in order to minimize service deterioration, fare increases and ridership losses through five public transport systems. As shown dramatically by the five case studies, the service improvements and fare structures for truly effective regional public transport would require substantial government subsidy. Meanwhile, fiscal austerity at every government level is leading to subsidy cutbacks in most countries of Europe and North America.

Allport (1981) compared the total costing of bus, light rail and metro public transport systems to define the economic costs of operations, maintenance and administration on a common basis. Based on 8 km radial transport corridor with realistic supply and demand characteristics, it was found that the bus had the lowest operating cost up to a travelling demand of 37500 passengers. The light rail transport was always the least cost in the range of 100000 passengers to 175000 passengers, and the metro was the only mode with a carrying capacity above this level. When travelling time value is taken into account, the bus is always the least-cost mode up to 50000 passengers. Guerra (2011) and Gomez-Ibanez (1985) pointed that light rail transport is expanding rapidly in developed cities with low corridor volumes, sometimes feeding heavy rail systems. While in developing countries, LRTs exist only in several larger cities. The cost of construction and operation management varies widely but it is considerably higher than the cost of alternative public transport form BRT.

Guerra and Cervero (2011) found that capital costs per route-kilometer of metro vary substantially between cities, between metro systems, and between metro lines within the same city and system. They are taxing even for developed nations: $50–$150 million per kilometer, with cost over-runs being the rule rather than the exception. Ben-Akiva and Morikawa (2002) pointed that in addition to high capital costs, metro systems have high operating costs and usually require subsidies; otherwise the price of the tickets would be prohibitive even in developed cities. While in principle public transport operations do not need to be profitable, the high capital and operation cost of metros makes them less economically viable in medium-sized developing cities than in megacities. Pagliara and Papa (2011) prompted that due high costs, developing cities often can only construct metro systems over a few kilometers in a few limited corridors, which do not meet the broader transport needs of the population. Additionally, the public sector may end up with a long-term debt that can affect investment in more pressing policy areas.

2.2.2 Internal factors on metro ridership

Studies of influential determinants on metro ridership from operators' perceptions and views emphasize internal factors, such as service improvements and adjustments, fare innovation and changes, marketing and information, new planning approaches and partnerships, service quality and coordination. Here literature of internal factors will be split into service quantity and service quality.

2.2.2.1 Service quantity factors

Derrible and Kennedy (2009) found that ridership is not solely determined by cultural characteristics (North American, European, Asian) or city design (transport-oriented, automobile-oriented), there is a significant relationship between network design and annual ridership per capita, achieving a goodness of fit of 0.725 in multiple regression analysis. They (2009) first developed three network indicators by studying 19 worldwide metro systems, which are termed as coverage, directness, connectivity; an application of this model can be found in Toronto metro systems (2010). Then, they developed another three indicators of system characteristics for 33 worldwide metro networks using this methodology, which are termed as state, form and structure (2010).

Boccaletti (2006) found that scale-free patterns and small-world effects have emerged to be particularly relevant in metro systems and pointed that network topologies play a key role in attracting people to use public transport. Angeloudis and Fisk (2006) pointed that the more expensive system infrastructure is, the more a transport system should tilt toward the hub-and-spoke concept. Levinson (2012) systematically compares a set of network structure variables across the 50 largest metropolitan areas in the United States: connectivity, hierarchy, circuity, treeness, entropy, accessibility. The results showed that a 1 percent increase in accessibility reduces average metropolitan commute time by about 90 seconds each way; a 1 percent increase in network connectivity reduces commute time by 0.1 percent, which is beneficial to organize passenger flow.

Vuchic and Musso (1991) pointed that alignment of network design sometimes deviates where needed to provide local access. Accessibility for metro user requires close stations, but the line-haul function requires high speed and therefore long distances between stations. Most metro lines provide access to/ from various points in the corridor they serve and transport passengers over longer distances along that corridor. If a metro system is intended to provide area coverage, station spacing along the line should be 500 to 800 m long. Long lines which serve suburbs of very large cities have average spacing between 1,000 and 3,000 m. Some cities use a combination of urban metros systems to serve central urban areas and regions, respectively. Long lengths of transport lines serve directly more trips than short lines and have a smaller proportion of terminal time, thus allowing better utilization of personnel and vehicles. On the negative side, long lines may result in less efficient scheduling as well as a problem of frequent delay propagation.

Demery (1994) pointed that demand-supply equilibrium as a determinant of transport ridership was not addressed during early planning for several recent fixed-guideway projects. His study outlines three principal parameters for service supply factors: maximum utilized capacity or vehicle occupancy (passengers per vehicle per hour); maximum service level (vehicles per hour) and the share of weekday ridership carried during the peaking hour in the busier direction. Walker (2015) pointed that frequency, distance and speed are related with the high-ridership transport product. Lee et al (2008) found that the weighted passenger distribution of 380 stations in the Metropolitan Seoul Metro system displayed a power-law behavior whereas the strength distribution follows a log-normal one by analyzing various network measurements, including path length, clustering coefficient, diameter, and radius as well as the efficiency of the network.

Jarrett Walker and Christopher Yuen (2018) pointed that the coverage goal is always the opposite of a ridership goal. Sung et al. (2014) prompted that a rail transport service coverage boundary of 500 m provides the best fit for estimating rail ridership levels. The results of a regression model confirmed that land use density is positively related to rail ridership within a 750 m radius of each station. In contrast, land use diversity is not associated with rail ridership. They also found that station-level accessibility is as important as land use for explaining rail ridership levels.

2.2.2.2 Service quality factors

Handy (2002) find that the quality of service, such as customer and on-street service and station and on-board safety, is more important in attracting riders than changes in fares or the quantity of services. Vuchic and Musso (1991) found that the experiences from the design and operation of some large older metro systems (such as London, Moscow, New York) or from numerous recently built medium-size metro systems (such as Hong Kong, San Francisco, Sao Paulo) remain largely unknown to the designers of new metro networks.

Bhandari et al. (2009) found that metro systems as public transport projects show a positive impact on equity of mobility and accessibility and also are examined to lead to a decrease in the generalized travelling cost by other modes. Angeloudis and Fisk (2006) promoted that the characteristic high connectivity but low maximum vertex degree of metro networks provides robustness to random attack. Zhang et al (2011) argued that metro systems are robust against random attacks but fragile for malicious attacks, the highest node-based attacks can cause the most serious damage among the different attack protocols. Two novel parameters called functionality loss and connectivity of metro lines are proposed in order to assess the reliability and robustness of Shanghai metro systems in China. Derrible and Kennedy (2010) found that most metro networks show atypical behaviors with increasing size. In order to increase the robustness of metro systems, they suggested that it is important to create additional transfers possibly at the periphery of city centers for larger networks and create transfer stations for smaller networks.

3. Methodology

3.1 Study areas

Several metro systems around the world were selected as study areas for analysis. In this study, metro means urban mass rapid rail transport with its own right-of-way, whether it is underground, at grade or elevated. A wide arena of metro systems was observed over six major world areas including Africa, Asia Europe, North America, South America, and Oceania, thereby considering different cultures and specificities. A pattern of developments of network systems might be achieved as well, via collecting metro systems of all sizes from 1 to 25 lines.

The metro systems used in this study is mainly from on metrobits.org (last updated 26 June 2018), which is an independent, non-profit organisation. The main database maintained by the organisation, World Metro Database, shows that there are 213 metro systems worldwide, 711 lines with a combined length of 14744 km and 12320 stations including 1128 transfer stations with 1.27 km average station distance. It is interesting to note that metro systems in Asia carry over 70 million daily ridership which represents nearly half the world total. Tokyo has the busiest metro system in Asia and also in the world, with close to 3.3 billion passenger trips per year. The metro system in Moscow carries over 2.4 billion per year, making it the busiest outside Asia (world's 3rd). New York City has the highest ridership in North America (1.7 billion, world's 7th) and Mexico City is the busiest network in Latin America (1.6 billion, world's 8th). The metro systems of Paris opened in 1900 and London opened in 1863 are ranked 9th and 11th respectively in the world.

A degree of caution should be exercised in the identification of study areas. Although the term Metro is used in most cities with metro systems around the world, the term

Subway is commonly used in America. Interestingly, Glasgow in Scotland officially has a Subway while urban rail transport systems in Washington and Los Angeles are named by Metro. The term Subway is also used in Japan and South Korea, although Metro is used in other parts of Asia. Many other cities or countries also have their own terms, for example Underground or Tube (in London), U-Bahn (in Germany and Austria), T-Bane (in Sweden and Norway), Subte (in Buenos Aires), MRT (in several Asian cities standing for mass rapid transit).

On the other hand, complex urban rail systems also make it controversial to differentiate metro systems from other rail transport systems. When a metro system has gradually evolved from former heavy trains, trams and monorails, it is ambiguous to measure the actual length or date of inauguration of that metro system. Besides, the differences between different modes of transport are fading since the middle of the 20th century as hybrid trains make it possible to integrate sections on streets with tunnel and/or heavy-rail sections.

3.2 Data collection

A total number of 213 global metro systems was provided by World Metro Database from the metrobits.org, an independent, non-profit organization. The list of metro systems from Wikipedia provided 160 metro systems from 157 cities in 55 countries and 42 metro systems currently under construction. In order to collect a relatively complete list of metro system worldwide up to 2017, the two datasets were emerged by the names of metro systems. Obviously, data provided by the two databases was collected from different years, and ridership data from World Metro Database was incomplete and varied from 2005 to 2012, while in the list from Wikipedia, ridership data mostly around 2016 and 2017 are not stable enough for analysis. Hence, basic information for each metro system in 2017, such as lines, length, stations and annual ridership, was mostly updated by each individual transport authority website, although a few websites were required to translate in English. Other data used in this study from the metrobits.org. were crowd-sourced by public transport enthusiasts including John Kennes, Nordpil, Mike Rohde, Jordi Serradell. The estimated urban population for the year 2017 was collected from United Nations Statistics as well. Notably, the one-year time period was based on the fiscal year rather than the calendar year.

In order to merge the list of metro systems with economic factors, income levels and development conditions are collected from two sources separately: The Identification for Development (ID4D) Global Dataset and Standard country or area codes of United Nations Statistics. The ID4D Dataset was compiled by the World Bank Group which provides a global estimate about development information split by country, region and income level. The Standard country or area codes was applied to get more specific and detailed geographic distribution information of metro systems. It contains the names of countries or areas in alphabetical order, three-digit numerical codes used for statistical processing purposes by the Statistics Division of the United Nations Secretariat and three-digit alphabetical codes assigned by the International Organization for Standardization (ISO).

The Global BRT Data consolidates data about bus priority systems in 168 cities worldwide from a variety of sources including researchers, transport agencies, municipalities and NGOs, which provides names of cities with BRT systems 2017. The List of tram and light rail transit systems from Wikipedia collects 398 global cities using tram, streetcar and light rail systems as regular public transit systems. The number of cities possessing light rail system is 62 around the world. All the collected date about other MRT systems was merged into the list of metro systems by the names of cities.

3.3 Variables used in the analysis

The variables first considered for analysis were chosen both through the investigation of past literature and examination of selected databases. The first factor to be determined was the dependent variable, annual ridership per capital by metro. Although both annual ridership and daily ridership were provided from the database, annual ridership was preferred to use since it possessed more evident number for analysis compared to daily ridership. This study formatted the dependent variable Ridership by annual metro ridership dividing city population, representing the number of unique journeys on a metro system every year provided for each local resident on average. It is important to note that by updating ridership data of various transport authority website, the sample would have a selection bias due to a discrepancy that some metro systems count transferring between lines as multiple journeys, but others do not. Besides, city population severed by metro systems might have a discrepancy with the collected city population provided by United Nations Statistical Databases. Table 1 includes all the variables with description.

Table 3.3.1 Variables with descriptions

Variable	Type	Description
Continent_name	Cat.	Mainland in which the metro system is located.
Continent	Cont.	Mainland in which the metro system is located quoted by Standard country or area codes in 2017 in six categories (Africa, Asia, North America, South America, Europe, Oceania, Antarctica)
Region	Cont.	Region in which the metro system is located quoted by ID4D Global Dataset in 2017 in six categories (South Asia Region, Africa Region, East Asia Pacific, Latin and Central America Region, East and Central Asia, Middle East & North Africa)
Country_name	Cat.	Sovereign state in which the metro system is located quoted by Standard country or area codes in 2017 in 120 categories.
Least_Developed_Countries	Cat.	Does a metro system belong to a least developed country? (yes/no)
Land_Locked_Developing_Countries	Cat.	Does a metro system belong to a least developed country? (yes/no)
Small_Island_Developing_States	Cat.	Does a metro system belong to a least developed country? (yes/no)
Development_level	Cont.	The development level of a country in which the metro system is located quoted by ID4D Global Dataset in 2017. (Developed countries, Developing countries)
Income_level	Cont.	Income Level of Sovereign state in which the metro system is located quoted by ID4D Global Dataset in 2017 in four categories (High income, Upper middle income, Lower middle income, Lower income)
City_name	Cat.	Primary city served by the metro system.

Population	Cont.	City population served by a metro system in million quoted by United Nations Statistics Division in 2017.
System_name	Cat.	Most common English name of the metro system.
Year_opened	Cont.	The year the metro system was opened for commercial service at metro standards.
Lines	Cont.	Number of lines in the metro network quoted by the system's operating company in 2017.
Stations	Cont.	Number of unique stations in the metro network quoted by the system's operating company in 2017. (Each interchange station is counted only once.)
Network_length	Cont.	The sum of the lengths of all unique routes in the rail network in kilometres quoted by the system's operating company in 2017. (Each route is counted only once.)
Annual_ridership	Cont.	Number of unique journeys on the metro system in million quoted by the system's operating company in 2017. (Each transferring journey is counted only once.)
Daily_ridership	Cont.	Number of unique journeys on the metro system every day in million.
Ridership_year	Cat.	The year of ridership collected ranging from 2005 to 2017.
Ridership	Cont.	Number of unique journeys on the metro system per person in million in 2017.
Stations_per_line	Cont.	Average stations per line in the rail network.
Ridership_per_km	Cont.	Number of unique journeys per kilometre on the metro system every year in million.
Average_line_length	Cont.	Average length per line in the rail network in kilometres.
Average_station_distance	Cont.	Average distance between two station in the rail network in metres.
Usage	Cont.	Efficiency rate used by residents.
Fare	Cont.	Prices of single a ticket for a journey of approximately 10 km or 10 stops.
Tracks	Cont.	The direction of tracks which train driving on the railroad in two categories (Right, Left)
Track_gauge	Cont.	Distance between the two rails forming a railway track.
Train_width	Cont.	Width per train in the rail network in miles.
Power_supply	Cont.	The methods of power supplied for the rail network in two categories (overhead wire, third rail)
Voltage	Cont.	The electricity volume supplied for the rail network in voltage.
Rubber_tyred	Cat.	Rubber-tyred trains used in a metro system. (yes/no)
Driverless	Cat.	Driverless technology available in a metro system. (yes/no)
Platform_screen_doors	Cat.	Platform screen doors available in a metro system. (yes/no)
Air_condition_trains	Cat.	Air condition used on metro train in a metro system. (yes/no)
Walkthrough	Cat.	Does it belong to a walkthrough metro system? (yes/no)
BRT	Cat.	BRT system available in a city with metro system (yes/no)
LR	Cat.	LR system available in a city with metro system (yes/no)

Sources: data from World Metro Database (2012); The list of metro systems from Wikipedia (2017); United Nations Statistics Division (2017); Identification for Development (ID4D) Global Dataset (2017); Standard country or area codes (2017); Global BRT Data (2017); The List of tram and light rail transit systems from Wikipedia (2017).

Collected data had to be restructured due to the form of text information from different sources, which were first transformed into continuous variables to reduce the overall number of considered independent variables and satisfy the requirement of correlation analysis for numerical data. Prior to the selection of independent variables, a correlation of all considered variables was performed (see Appendix A). It is not only useful to test the relative association of each independent variable on the dependent variables, but also to detect if any two independent variables were too highly related to each other. In order to analyse multiple influencing factors from a global view, all considered variables were then restructured into six main factors ranking.

Geographical factors were considered from continental level in which a metro system was located and formatted as Continent variable. From the summary of the collected data, there were 48 metro systems in Eastern Asia, 16 in Eastern Europe, 16 in Latin America and the Caribbean, 16 in Northern America, 15 in Western Europe, 13 in Southern Asia and 37 in Other part of the world. Subsequently, a scatter plot of Continent and ridership per capital was run to inspect the relationship of linearity (see Appendix B). The plot displayed no definitive evidence of a nonlinear relationship and thus the linear form was maintained for this study. In the same manner, Income, City, Form, Years and MRT were checked for linearity as well.

Economic factors were considered from country level in which a metro system belonged to and formatted as Income. Development level and income level were compared to determine which would be better for analysis. Although development level had lightly higher correlation with ridership per person than income level (0.45 vs 0.40), income level was selected for analysis given that it provided more detailed economical information. Subsequently, a scatter plot of Income and ridership per person displayed no definitive evidence of a nonlinear relationship and thus the linear form was maintained.

Demographic factors were considered from city level in which a metro system provides public transport service. A scatter plot of Population and ridership per person displayed no definitive evidence of a nonlinear relationship and thus the linear form was maintained. In order to inspect the influence of city types on metro ridership per capital, the continuous variable Population was converted to a categorical variable City based on the criteria of demography, grouping these data into five categories as Megalopolis with the population over 10 million, Conurbation with the population around 3 to 10 million, Metropolis around 1 to 10 million, Large city around 0.3 to 1 million and City around 0.1 to 0.3 million.

Another three factors selected from system level were related to system accessibility, operating experience and other mass rapid transport modes, and formatted as Form, Years and MRT variable respectively. When selecting suitable explanatory variables from World Metro Database, it is obvious that Lines, Stations, Network length, Length per resident, Average line length, Stations per line and Average station distance have higher correlations between each other., which should be excluded first

Form is introduced to estimate system accessibility and illustrate how metro systems are integrated in the built environment. System accessibility here is the accessibility of a metro system at the regional level, rather than physical accessibility by transport user. The average line length and the number of stations are usually used to identify whether a metro system is regionally or locally focused. Generally, longer average line length mean that metro systems can reach further out in the suburbs representing regionally-oriented accessibility, while metro systems with small average line length and many stations would increase local coverage. Three typical attributes of metro networks: route length, number of lines and number of stations are collated from the database to construct the factor Form. Referencing the literature (Derrible and Kennedy, 2010), metro systems are plotted by the average line length and the number of stations (see Figure 3.3.1), there is an obvious threshold, which can be roughly estimated as (150, 20), to group metro systems into three categories coded as regional accessibly, regional coverage and local coverage. Regional Accessibility refers that metro systems focus on connecting people in the outer city to the city core, allowing the suburban population to use the metro. Local Coverage refers that metro systems focus on servicing the city core, making metro a prime transport mode in downtown. Regional Coverage refers to a mix of the other two metro systems, generally well serving people in the city core and connecting people living in the surrounding regions. From the Figure 3.3.1, metro systems such as the Washington DC and St Petersburg metros emphasized on connecting people in city outskirt to city centre. Other cities such as Paris and Barcelona have extensive metro networks in city core. Metro systems in London, Tokyo and Moscow are seeking to do both. In this study, the same threshold was used to divide all the metro systems we collated for the year 2017 into the same three categories for later analysis.

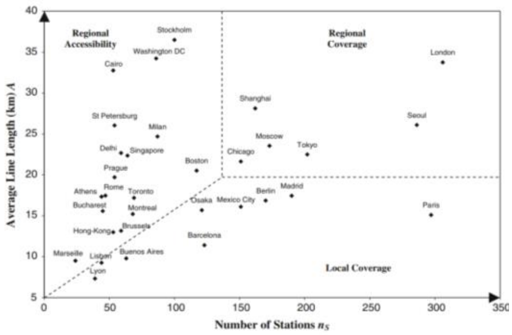


Figure 3.3.1 Form of metro systems

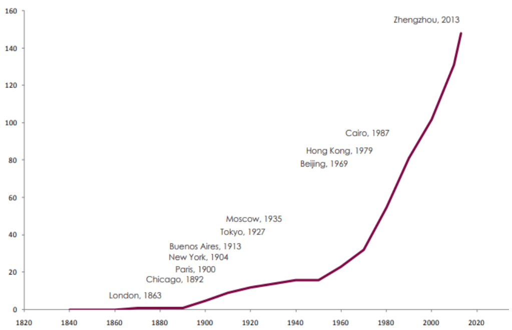


Figure 3.3.2 Number of metro cities and some key opening dates

Two typical MRT modes, BRT and LR, were picked up to analyse whether the presence of other MRT modes would affect personal metro ridership. The binary variable BRT and LR were recorded as dichotomous value of whether a metro system existed with BRT system in a city at the same time or LR system in a city at the same time separately. However, there were some overlays that some cities own three types of modes meanwhile since there was no definite limit for a city to develop multimodal transport. An adjustment is required to make in order to maintain consistency and thoroughness period to analysis. The two variables of dichotomous values were emerged to a categories variable regarding whether a BRT or LR existed in a city with metro in four categories as neither BRT system nor LR system, only BRT system, only LR system, both BRT system and LR system.

In order to inspect the influence of operating experience on ridership per capital, the age of a system was obtained by 2018 minus the opening year of each metro system, termed as Years. Figure 3.3.2 shows the growth trend of metro cities in number with some key opening dates. From the graph of years and number of systems, it can be seen that firstly, the construction of metro system starts in 1863 in London, increase obviously after 2000. Subsequently, a scatter plot of Years and ridership per person displayed no definitive evidence of a nonlinear relationship and thus the linear form was maintained.

Additionally, a number of variables had to be removed while cleaning datasets for the regression models. As a substantial number of variables in World Metro Database were not available for all the metro systems which are finally selected to analysis, they were therefore removed in order to avoid decreasing sample size further. This results in the removal of Usage, Fare Tracks, Gauge, Train width, Power supply and Voltage. Also, Rubber-tyred, Driverless, Platform screen doors, Air condition trains and Walkthrough was unable to be included due to not being collected prior to 2017.

A summary of the final variables for the regression models displays in Table 3.3.2. Main variables in the sample of the final cleaned datasets (i.e. NAs removed) was evaluated for comparison. It is immediately apparent that Asian systems constitute a much larger share of systems considered compared to those located elsewhere (48.13%). The percentage of metro systems belonging to high income level country is much higher for those belonging to lower middle income level country (50.63% vs 11.25%). Also of interest is that Metropolis with 1 to 3 million population constitute a much larger share of systems considered than other types of cities in the sample (40%). Regarding other

MRT modes, 35% metro systems exist with BRT or LR systems in a city meanwhile. The 85.63 percent of metro systems in the sample provide regional accessibility for residents. The average age of metro systems ranging from 1 year to 128 years is 34.49 years.

Table 3.3.2 Final variables with descriptions

Variable	**Description**	**Type**	**Values**	**Number of systems**
Continent	Where the metro system located	Cat.	Asia(base)	77 (48.13%)
			Europe	50 (31.25%)
			Latin America and the Caribbean	16 (10.00%)
			Northern America	15 (9.38%)
			Africa	2 (1.25%)
Income	Average income level in a country	Cat.	High income level (base)	81 (50.63%)
			Upper middle income level	60 (37.50%)
			Lower middle income level	18 (11.25%)
			Lower income level	1 (0.63%)
City	Type of city according to city population	Cat.	Megalopolis over 10 million (base)	19 (11.88%)
			Conurbation with 3 to 10 million	40 (25.00%)
			Metropolis with 1 to 3 million	64 (40.00%)
			Large city with 0.3 to 1 million	32 (20.00%)
			City with 0.1 to 0.3 million	5 (3.13%)
MRT	Is there any other MRT modes (i.e. BRT system and LR system) available in a city with metro system?	Cat.	Neither BRT nor LR available (base)	104 (65.00%)
			Having BRT system only	32 (20.00%)
			Having LR system only	17 (10.63%)
			Both BRT and LR available	7 (4.38%)
Form	Type of system accessibility	Cat.	Regional accessibility (base)	137 (85.63%)
			Regional coverage	15 (9.38%)
			Local coverage	8 (5.00%)
Years	How long has a metro system existed in a city?	Cont.	1 year to 128 years	34.49

*** Significant at 0.01 level, ** Significant at 0.05 level, * Significant at 0.10 level.

3.4 Analytical model

The Multiple linear regression is a statistical model that creates a "line of best fit" for independent variables to predict the outcome of a dependent variable. In the case of this study, regression models will be used to inspect the relationship between the chosen independent variables in Table 3.3.2 and metro ridership per capital. The linear relationship for a regression model is the basis for predicting the outcome of metro ridership per person and expressed as follow:

$$Y_i = \alpha + \beta_i X_i + \varepsilon_i \quad \text{for } i = 1, \dots, n$$

(Formula 3.4.1)

When Y_i is the dependent variable, α is the intercept (constant), β_i is the coefficient for each predictors X_i, and ε_i is the error (deviation) with i representing the ith respondent of sample. The coefficient β_i indicates the slope or gradient for each predictor in the regression model showing the relative association with the dependent variables holding all other variables constant. The coefficient also has a corresponding value, the t-statistic, which is used to test the level of significant of each predictors. The t-statistic is derived from testing the null hypothesis that a coefficient is zero for a given predictor ($H_O: \beta = 0$). If the null hypothesis is false at a 0.05 level of significance, then the coefficient is said to be able to significantly predict the dependent variables. However, due to the small sample size, it can be difficult to find factors significant at the 0.05 level and thus this study will recognize relationships as to have influential correlation if significant at the 0.10 level.

When performing regression, a number of assumptions are made concerning the errors are independent and normally distributed with a mean of zero and variance of σ^2. The latter assumption can be incorporated into Formula 3.4.1 for this study. The linear relationship for a regression model can also be written as:

$$Y_i \sim N(\alpha + \beta_i X_i, \sigma^2) \text{ for } i = 1, \dots, n$$

(Formula 3.4.2)

Where X_i represents diverse factors including Continent, Country, City, Form, Years and MRT (see Table 3.3.2). It is important to note that Y_i needs to follow a normal distribution, and the Tukey's Ladder of Powers transformation of ridership per person was required in order to perform the regression model, since values for the unaltered variable were greatly skewed towards lower ridership per capital. Hence, when interpreting the coefficients (β_i), they will now be based on Tukey's Ladder of Powers transformation of ridership per person, meaning that values will be read as the unit change of the independent variables will result in a 0.175 percentage change in ridership per capital while holding all other variables constant.

The model fit of the regression models was analysed through the estimation method of ordinary least squares (OLS). OLS calculated the deviation between squares of the observed, predicted and mean values of the model. The three resulting variables calculated are the total sum of squares (TSS), which is the sum of the error sum of squares (ESS) and the regression model sum of squares (RSS). These values can be expressed in the following manner:

$$\sum_{i=1}^{n}(Y_i - \bar{Y}_i)^2 = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 + \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y}_i)^2$$

(Formula 3.4.3)

Where the TSS is the total squared difference from the observed values (Y_i)and mean values ($\bar{Y}_i$), the ESS is the total squared difference of the observed values (Y_i) and predicted values ($\hat{Y}_i$), and the RSS is the total squared difference between the predicted values ($\hat{Y}_i$) and mean values ($\bar{Y}_i$). These values are then used to estimate the F-value and r^2 to assess the model fit._The F-value acts in a manner similar to the t-statistic in that a null hypothesis that all model coefficients have a value of zero ($H_0: \beta = 0$) is tested. If this is found to be false at a significant level, the null hypothesis is then rejected meaning at least one of the predictors in the model significantly predicts a relationship. The r^2, which is the proportion of the variance that can be explained by the regression model, was utilised to ascertain the best fit to describe the dependent variables. The r^2 being a product of the estimation method of ordinary least squares is expressed as follows:

$$r^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y}_i)^2}$$

(Formula 3.4.4)

Where it can be seen that r^2 is one minus the ESS divided by the TSS. As the r^2 will increase for each variable that is added to a regression model, the adjusted r^2 was referenced when checking model fit as it penalises each additional independent variable that is added to a model and can be written as:

$$Adjusted\, r^2 = 1 - (1 - r^2) * \frac{n-1}{n-k-1}$$

(Formula 3.4.5)

In which n is the sample and k is the number of independent variables in the model. The resulting value is a ratio and as such represents the percentage of variation accounted by the model. These factors will be discussed when interpreting the results of the model in section 4.

3.5 Model diagnostics and adjustments

A number of diagnostics and assumptions were carried out to ensure the generalizability of the model (see Appendix B). The assumptions, linearity of variables and independence of errors were found to be met for the primary regression model based on 160 metro systems in general.

However, a few assumptions warranted more in depth explanations as they required additional inspection. When inspecting for homoscedasticity, the constant variance of error, the residuals were observed in the residual versus fitted value graph. It appears homoscedastic, and the consequent of non-constant error test (NCV) revealed that the model violated the null hypothesis of homoscedasticity ($p < 0.05$). The presence of multi-collinearity was inspected through correlation analysis as well as a variance inflation factors test (VIF). The VIF test revealed no values above 10, which is considered a threshold of concern (Myers, 1990). The independence of errors was inspected through Durbin Watson test, the results showed the model did not violate the null hypothesis of homoscedasticity ($p > 0.05$). Lastly, to ascertain the model to best represent the sample, Cook's distance was used with a cut-off (0.025) as it has been reported that under this threshold, an observation does not have a large effect on the regression analysis (Stevens, 2009). One Outliers detected is the metro system in Glasgow due to the value of Cook's distance (0.075) and studentized residuals (-3.815268) with the consequent of Bonferonni test ($p < 0.05$). Thus a metro system in Glasgow as one study area was deleted, the same process of diagnostics and assumptions for the remaining data was carried out again and all the assumptions were found to be met for the final regression model based on 159 metro systems.

4. Results

This section examines the relationship between six groups of selected variables and annual ridership per capital of 159 global metro systems. A multiple linear regression model has been set up to explain how these factors are potentially related to metro ridership and the corresponding consequent effect on metro systems design and planning. The final results are presented in Table 4.1 below.

Table 4.1 Final results of the regression model

		95% confident interval		
Variables	**Coef.(*β*)**	**Lower**	**Upper**	***t*-statistic**
(Intercept)	1.1126	0.7440	1.4812	5.968***
Continent (Reference: Asia)				
Europe	-0.0077	-0.2076	0.1921	-0.076
South America	-0.0216	-0.2717	0.2286	-0.170
North America	-0.3431	-0.6171	-0.0690	-2.474*
Africa	0.0044	-0.5317	0.5404	0.016
Income (Reference: High income level)				
Upper middle income level	-0.0525	-0.2397	0.1347	-0.555
Lower middle income level	-0.2567	-0.4786	-0.0348	-2.287*
Lower income level	-0.8264	-1.5780	-0.0748	-2.174*
City (Reference: Megalopolis over 10 million)				
Conurbation around 3 to 10 million	0.0082	-0.2227	0.2392	0.070
Metropolis around 1 to 3 million	0.0107	-0.2325	0.2539	0.087
Large city around 0.3 to 1 million	0.2698	-0.0279	0.5676	1.792#
City around 0.1 to 0.3 million	0.6897	0.2337	0.1458	2.990**
Other MRT (Reference: Neither BRT nor LR available)				
Only BRT system	0.1210	-0.0469	0.2888	1.452
Only LR system	0.0323	-0.1822	0.2467	0.297
Both BRT and LR available	-0.2901	-0.6133	0.0331	-1.775#
Form (Reference: Regional accessibility)				
Regional coverage	0.2500	-0.0014	0.5014	1.966#
Local coverage	0.2384	-0.0619	0.5387	1.569
Years	0.3928	0.2981	0.4875	8.198***
F-value (p-level)	14.21 (< 2.2e-16)			
Adjusted r^2	0.5871			

*** Significant at 0.01 level, ** Significant at 0.01 level, * Significant at 0.05 level, # Significant at 0.1 level.

Overall, the F-value for the model is false at a significant level ($p < 0.05$), thereby the null hypothesis that all model coefficients have a value of zero is rejected, which means at least one of the predictors in the model significantly predicts a relationship. The $Adjusted\ r^2$ value of the regression model is 0.5871, which accounts for the 58.71 percent variation in metro ridership per capital and ascertain a better model fit.

Continent is introduced as a group of variables in five categories to describe system locations with specific continents. The reference category is metro systems located in Asia. The coefficients for those metro systems constructed in Europe, South America, Africa are insignificant. These dummy variables have a substantially reduced sample size, which may partly explain the lake of significant. However, the coefficient for those systems located in North America is negative and significant at the 5 percent level of significance. This indicates that metro systems located in North America are expected to have 34.31 percent (i.e., exp(0.3431)) less ridership per capital compared to metro systems in Asian areas, holding other groups of variables constant. Metro systems in Africa are the only variable to make just 0.44 percent more ridership per capital compared with the reference variable of Asian systems although it is not statistically significant, the coefficients direction of other continents all show a decreasing pattern.

Turning attention to the influence on personnel metro ridership from the perspective of country, average country income levels are considered as another group of explanatory variables termed as Income in four categories. The reference variable is High income level. All three coefficients attached to the dummy variables representing Upper middle income level, Lower middle income level and Lower income level are negative, with two of them also being significant at the 5 percent level of significance. More specifically, metro systems belonging to Lower middle income level country are expected to make 25.66 percent less ridership per capital compared to metro systems belonging to High income level country, holding other groups of variables constant.

Metro systems belonging to Lower income level country are expected to make 82.63 percent less ridership per capital compared to metro systems belonging to High income level country, holding other groups of variables constant.

The model also has been set up with the City factor included as an explanatory variable by grouping city population served by metro systems in five categories. The reference category is metro systems built in megalopolis with a population over 10 million. All four coefficients attached to the dummy variables representing conurbation, metropolis, large city and city are positive, with two of them also being significant at the 10 percent level of significance and 1 percent level of significance separately. These indicate that metro systems built in large cities with a population around 0.3 to 1 million are expected to make 26.98 percent more ridership per capital compared to metro systems severing Megalopolis, holding other groups of variables constant; and metro systems built in cities with a population around 0.1 to 0.3 million are expected to make 68.97 percent more ridership per capital compared to megalopolis, holding other groups of variables constant.

Another group of variables MRT is introduced to inspect the influence of the presence of other MRT modes on metro ridership per capital. The reference category is cities with only metro systems. The coefficients for those cities with another MRT mode, either BRT or LR, are positive but insignificant. These may suggest that a metro system existing with a BRT system are expected to make 12.09 percent more ridership per capital compared to only a metro system in a city, holding other groups of variables constant, while metro systems existing with LR systems are expected to make just 3.22 percent more ridership per capital, although these results are not statistically significant at 5 percent level. The coefficient for those cities with multimodal transport options, i.e. where both LRT and BRT systems are available in a city, is negative and significant at the 10 percent level. This indicates that cities with multimodal transport options are expected to make 29.01 percent less ridership per capital compared to cities with only metro systems, holding other groups of variables constant.

Form is used to measure the influence of system accessibility on metro ridership per person and introduced as a group of variables in three categories. Larger service coverage provided by a metro may attract more commuters travelling by this mode. The reference category is metro systems with Regional accessibility. Both coefficients of dummy variables representing Regional coverage and Local coverage are positive, only metro systems with Regional coverage is found to be significant at the 10 percent level of significance, which are excepted to have 24.99% more ridership per person than the reference group of Regional accessibility, holding other variables constant.

The opening years of 149 metro systems are collected as an explanatory variable and termed as Years. The coefficient for Years is positive and significant at the 1 percent level of significance. The age of systems is also the most significant factors on metro ridership per capital from the regression model. This indicates that a one-unit increase in years will increase 39.27 percent change in ridership per capital when holding other groups of variables constant.

5. Discussion

This section begins with a discussion of the results of the multiple linear regression model with reference to past literature. Generally, the findings are mixed compared to previous studies. Four external factors and two internal factors are discussed in order to understand how these factors are potentially related to metro ridership and the corresponding consequent effects on metro systems design and planning. Afterwards, implications in terms of existing metro systems will be present for local planners and agencies as well as to inform international institutions like the World Bank in their decision-making and target-testing process.

5.1 Discussion

Generally, the findings from the regression results are mixed compared to previous studies. According to the statistic results of global metro figures in 2014 (UITP, 2015), metro systems in Asia have carries over 70 million passengers per day which represents nearly half the world total. Based on the latest collected data of 2017, we choose Asia as the reference group to compare with other continents in this study. The four dummy variables measuring Continent are of interest here. Our argument is that location with specific continent may influence ridership per capital. It could also be argued that high ridership by metro is more likely to incur the construction of metro systems in a particular continent. Here we do not fully address the possibly of endogeneity and instead interpret these results as association and not causal effects. Compared to metro systems in Asia, we found that personal metro ridership significantly decreased in North America, controlling country's economic activity level, city population, system accessibility, age of system, the presence of BRT or LR systems constant. This is in line with the preview research by Pucher (1988) that there is negligible ridership growth in the proportion of travelling trips by mass public transport in the United States even if supported by large public transport subsidies. The metros managers in North America could take some actions to encourage the public use metros. It is interesting to note that the change of personal metro ridership in Africa, Europe and South America is extremely small compared to Asia, and metro ridership per capital only slightly increased in Africa compared to Asia, although they are not statistically significant.

Investment of metros is seen as a more permanent commitment of government to public. Supporting the previous argument promoted by Pojani and Stead (2015) that metros are generally more appealing to middle class passengers, our argument is that higher income levels would attract more residents travelling by metro. The reference variable is High income level. The results suggest that metro systems constructed in countries, where the average income level of residents is relatively lower, have less ridership compared to those systems belong to countries where the living standard of people are in better level. On another hand, from the previous literature that metro systems can improve social equity in terms of public transport service, however our results suggest that more ridership produced by metro systems are served for people in higher income level countries. Metro ridership taken by the poor seems extremely lower compared to people with high income, people with lower income may not often employ this public transport mode to satisfy their daily commuting. The implementation of metro projects is usually under government control given that relatively high capital and operational costs, therefore, it should take into careful consideration with construction project of metro network if the average income level of a country possesses in lower middle level, especially for lower income level courtiers, it is not a wise choice to adopt such expensive public transport projects due to comparatively lower ridership per capital.

Given that implement of a metro system usually requires a city with a population of 2-3 million having at least one transport corridor. Our argument is that cities accommodating larger residents may improve the use efficiency of metro systems, i.e. incurring more ridership per capital by metro. The reference category is metro systems built in megalopolis. The coefficients of metro systems built in larger cities with a population around 0.3 to 1 million and cities with a population around 0.1 to 0.3 million are positive and significant. It is interesting to note that the relationship appears to be monotonic that cities with lower population have more ridership compared to megalopolis when controlling other groups of variables constant. These results in particular are not consistent with our argument based on previous literature. An explainable reason might be that for medium-size cities with the population under 1 million, metro systems would become the dominant mode of public transport, thereby the maximum utility would be achieved, i.e. high ridership per capital.

The influence of multi-modal transport options on metro ridership is consistent with the previous literature that once a particular MRT system has developed, resources tend to be devoted to that system only, although theoretically cities should follow a balances approach by using complementary MRT systems appropriate to local circumstance (Wright and Fjellstrom, 2003). There may be a competitive or complemental relationship between other mass rapid transport modes and metros, our argument is that the presence of other MRT modes may influence metro ridership per capital. The factor MRT is introduced as group of variables in four types of neither BRT system nor LR system, only BRT system, only LR system, both BRT system and LR system. The reference category is cities with only metro systems. The results suggest that there is a complementary relationship between metro systems and BRT or LR, and when metros are limited to satisfy daily travelling demand, BRT seems a better choice than LR in terms of ridership per person, when controlling other factors constant, although they are not statistically significant from this model. There is a significant substitution effect of more multi-modal transport options on public transport ridership by metro. For a city demands to develop mass rapid transport modes, especially metros systems, the prediction of future passenger flow is questionable given that urbanization and the growth of global population. This has some potentially important policy implications, for example, Litman (2017) suggested that multimodal planning in urban transport is beneficial to satisfy the diversity of travel demands. However, our results suggest that cities exiting with other mass rapid systems will undermine metro systems from the terms of ridership. Multi-modal transport options should be adopted with careful consideration.

Given that the ridership goal of a particular transport system usually conflict with its coverage goal (Jarrett Walker and Christopher Yuen, 2018), our argument is that the accessibility of metro systems would influence ridership per capital, i.e. larger service coverage of metro systems may attract more residents travelling by this mode. The factor Form is introduced to measure the system accessibility of metros. The reference category is metro systems with regional accessibility. It is interesting to note that the influence of metro systems with regional coverage and local coverage on personal ridership are similar and positive compared to systems with regional accessibility, although only systems with Regional coverage is statistically significant compared to those with regional accessibility, holding other variables constant. Regional Coverage refers to a mix of the other two metro systems, generally well providing a comprehensive level of service in a defined part of the city and connecting people form the outer city to the city core.

The results for age of systems is consistent with our argument that the more mature systems are, the greater ridership per capita they produced. The coefficient for the operating years of metro systems is positive and significant at the 1 percent level of significance. It is also the most significant factors on metro ridership per capital of the overall pattern emerging from the results. Given that metro systems are often designed within a time-limited horizon and adjustments and correlations are also hard to make on the fixed railways and stations once construction completed, the influence of age of systems on metro ridership would be useful to take into consideration to predict the passenger flow before construction.

5.2 Policy implications

The discussion of these results offers a means to compare metro systems worldwide in order to help local planners and agencies in their decision-making and target-testing process, implications in terms of existing metro systems will be present for international institutions such as the World Bank, and other investment banks such as Inter-American Development Bank, Asian Development Bank etc. Figure 5.2.1 shows the network maps of six bustiest metro systems around the world. They are London Underground, New York City Subway, Paris Metro, the MTR Hong Kong, Shanghai Metro and Moscow Metro separately from left to right.

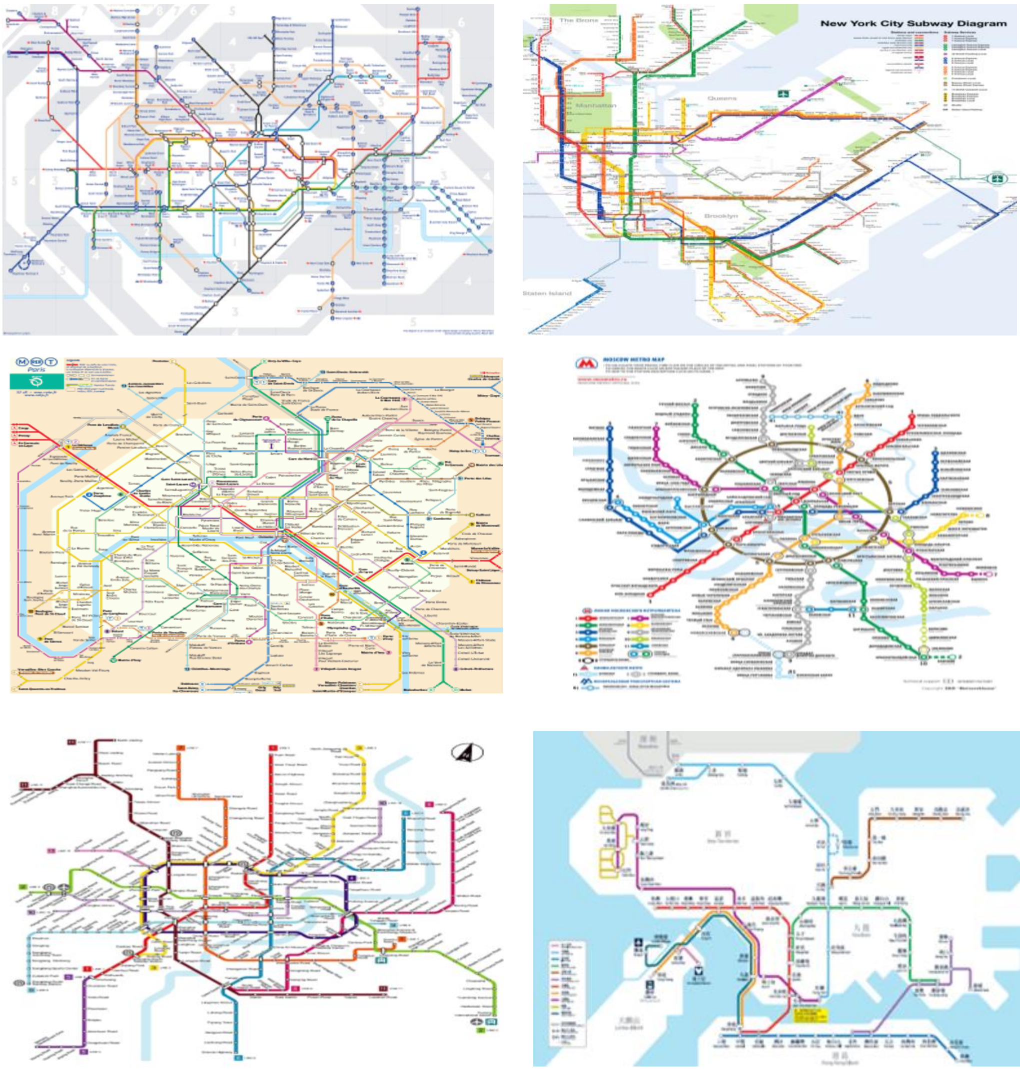


Figure 5.2.1 Network maps of six bustiest metro systems around the world

Source: World Metro Database

According to the result that personal metro ridership significantly decreased in North America compared to Asia, transport policy, such as encouraging local residents to use public transport instead of private, would be recommended in North America. Nearly two thirds of these networks are located in Asia and Europe, with the number of 54 and 46 respectively, 18 systems in Latin America, 16 in North America and 7 in the Middle East and North Africa region. Hence, some metro projects could be taken into consideration in some developed cities in North America as well. Generally, more ridership produced by metro systems are served for people in higher income level countries when controlling other factors. Therefore, it should take into careful consideration with construction project of metro network if the average income level of a country possesses in lower middle level, especially for lower income level courtiers, it is not a wise choice to adopt such expensive public transport projects due to comparatively lower ridership per capital. For some medium-sized cities with a population around 1 million, metro ridership per capital even high than megacities, when controlling other factors. Metros are an optional public transport mode for several larger cities but not as gigantic as megacities, which are usually designed around 30,000 to 40,000 passengers per hour in the peak direction. Additionally, multi-modal transport options should be adopted with careful consideration for a city which prefers to develop mass rapid transport, especially metros systems. Cities with three types of mass rapid systems will undermine metro systems in the terms of ridership, but there is a complementary relationship between metro systems and BRT or LR, and when metros are limited to satisfy daily travelling demand, BRT seems a better choice than LR in terms of ridership per person. Policy implication of the influence of system accessibility on metro ridership could focus on metro networks with regional coverage during the project planning phase. Enhancing regional accessibility favors the development of a few corridors that can strengthen the city core. The regional accessibility of metro systems can be achieved by increasing the average line length and the distance between the metro stations. However, having long lines reaching to suburbs can also result in urban sprawl. Metro systems with regional coverage generally well provide a comprehensive level of service in a defined part of the city and connect people form the outer city to the city core. For a metro system with lower ridership we can hold positive attitude given that personal ridership would increase with years significantly. This is in line with exceptions given that the more mature systems are, the greater ridership per capita they produced. We can also predict future values and trends of ridership per capital for a city with intention to construct metro system. If the ridership is too high, we could consider to introduce BRT or LR modes.

6. Conclusion

The gradual shift towards the adoption of rail-based public transport systems has been fuelled by the rapid pace of urbanization. Metros as a mass rapid transport mode have increasingly become viable choice for public transport development, especially in larger cities with high population densities. To alleviate the deficit of metro systems globally, it is imperative to understand the condition of operation efficiency in terms of metro ridership. The objective of this study is to examine factors that influence ridership of metro systems, with the goal of supplementing planning and policy decisions. Six main factors have been selected to investigate the influence on ridership per capital by reviewing past literature of influential factors on metro ridership in this study. Six groups of variables are related to systems locations with specific continent, country's economic activity level, city population, the presence of BRT or LR modes, system accessibility and age of system from two perspectives of external factors and internal factors. A wide arena of 159 global metro systems has been observed over six major world areas including all network sizes from 1 to 25 lines, thereby considering different cultures and specificities as well as the patterns of developments of network systems. The overall purpose of the study is to understand how these factors are potentially related to metro ridership and the corresponding consequent effect on metro systems design and planning. This study offers a means to compare metro systems worldwide in order to help local planners and agencies in their decision-making and target-testing process, as well as to inform international institutions such as the World Bank, and other investment banks such as Inter-American Development Bank, Asian Development Bank etc.

A multiple linear regression model then has been set up to understand how six selected factors explain the variance in annual ridership per person for the year 2017, thereby achieving a comprehensive and comparative understanding of metro systems. Continent is introduced as a group of explanatory variables in five categories to describe system locations with specific continent. Income is introduced as a group of variables in four categories to estimate the influence of the country income levels on metro ridership per capital. City is introduced as a group of variables in five categories of Megalopolis, Conurbation, Metropolis, Large city, City based on city population. Another group of variables MRT is introduced in four types of neither BRT system nor LR system, only BRT system, only LR system, both BRT system and LR system. Form is introduced as a group of variables in three categories to measure the influence of system accessibility on personal metro ridership. The opening years of 159 metro systems are collected as an explanatory variable as well.

According to our results that personal metro ridership significantly decreased in North America compared to Asia, holding other groups of variables constant. Metro systems belonging to those countries in lower middle income level are expected to make 25.66 percent less ridership than those countries in high income level, holding other groups of variables constant, and for some systems located in lower income level country, they are expected to make 82.63 percent less ridership per capital compared to high income level country, holding other groups of variables constant. Large cities with a population around 0.3 to 1 million are expected to make 26.98 percent more metro ridership per capital compared to megalopolis with a population over 10 million, holding other groups of variables constant. In some cities with a population around 0.1 to 0.3 million, metro systems are expected to make 68.97 percent more ridership per capital compared to megalopolis, holding other groups of variables constant. Cities with multimodal transport modes are expected to make 29.01 percent less ridership per capital compared to cities with metro systems only, holding other groups of variables constant. Moreover, metro systems with regional coverage are excepted to have 24.99 percent more ridership per capital compared to the reference category of Regional accessibility, holding other variables constant. A one-unit increase in Years will increase 39.27 percent change in ridership per capital when holding other groups of variables constant.

Some potential methodological issues in descriptive analysis could not be neglected, for instance, highly subjective data, biases based on limited or incorrect information, data collection process often not outlines in detail, no questions about perceived causality, questionable causal linkage. On other hand, for causal analysis, generalizability is limited due to mostly small sample sizes; problem with multi-collinearity between individual variables; endogeneity problems between service supply variables and demand; many promising variables are not included in models and some variables are hard to quantify.

In order to enhance the results of this study, some guideline for future research can be suggested. Six selected factors can be investigated in depth. This aspect has not been directly addressed in this study due to the lack of data; further research could therefor take more detailed variables in account by this model and other regression models like logistic regression. Meanwhile, thoughtful and reasonable variables selection could be paid more attention. The results of the continental effect on metro ridership per capital could be improved by controlling other influential factors from the same perspective in order to strengthen the interpretability and comparability of results. The income level of country has only been considered in this study, however, there are plenty variables of economic factors can be tested further to identify the association with metro use according to previous literature review. The classification of cites based on population might be also improved by referencing other demographic criteria or economical standards. Regarding to other mass rapid transport modes, we can also set up ridership models separately to compare the difference between them. We made up the Form factors to measure the system accessibility by referencing a literature, while the system accessibility might be measured better by using other specific methods or models based on graph theory. The age of systems in this study was used to represent the operational experience of system, given that there are many other significant factors associated with operational experience like labour education level and technology improvement, we could also inspect the influence of these factors on metro ridership per capital.

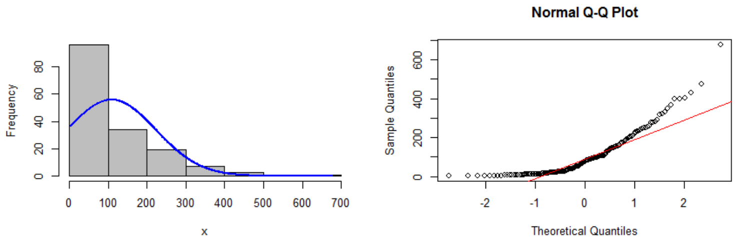
Appendix A: Correlation Table

	Ridership	Continent	Development_level	Income_level	Population	Lines	Stations	Network_length	Station_per_line	Ridership_per_km	Length_per_resident	Average_line_length	Average_station_distance	Opening_years	BRT	LR
Ridership	-															
Continent	0.16	-														
Development_level	0.45	0.39	-													
Income_level	-0.40	-0.27	-0.65	-												
Population	-0.15	-0.26	-0.48	0.35	-											
Lines	0.48	0.01	0.06	-0.16	0.44	-										
Stations	0.47	-0.04	0.02	-0.16	0.50	0.92	-									
Network_length	0.33	-0.09	-0.09	-0.06	0.62	0.87	0.91	-								
Station_per_line	0.12	-0.14	-0.19	-0.07	0.32	0.11	0.38	0.31	-							
Ridership_per_km	0.43	0.03	-0.01	0.07	0.35	0.34	0.32	0.26	0.04	-						
Length_per_resident	0.56	0.34	0.47	-0.42	-0.31	0.16	0.15	0.13	0.11	-0.16	-					
Average_line_length	-0.05	-0.10	-0.34	0.08	0.51	0.18	0.37	0.51	0.74	-0.05	0.13	-				
Average_station_distance	-0.16	0.08	-0.22	0.12	0.28	0.13	0.08	0.35	-0.07	-0.14	0.15	0.57	-			
Opening_years	-0.55	-0.35	-0.55	0.37	0.17	-0.48	0.45	-0.28	-0.03	-0.25	-0.33	0.12	0.16	-		
BRT	0.12	0.16	-0.20	0.02	0.19	0.23	0.19	0.17	0.02	0.11	-0.02	0.10	0.11	-0.01	-	
LR	0.17	0.20	0.20	-0.15	-0.04	0.22	0.16	0.17	-0.05	0.09	0.20	0.01	0.12	-0.33	0.08	-

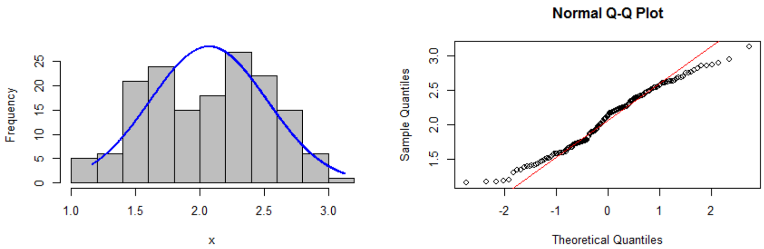
Appendix B: Model Diagnostics and Assumption

Primary Model - Model 1

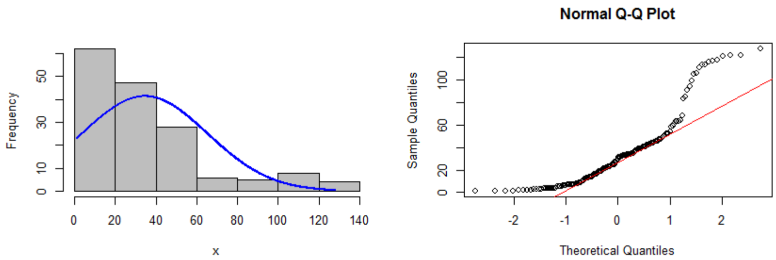
The distribution of the dependent variable Ridership



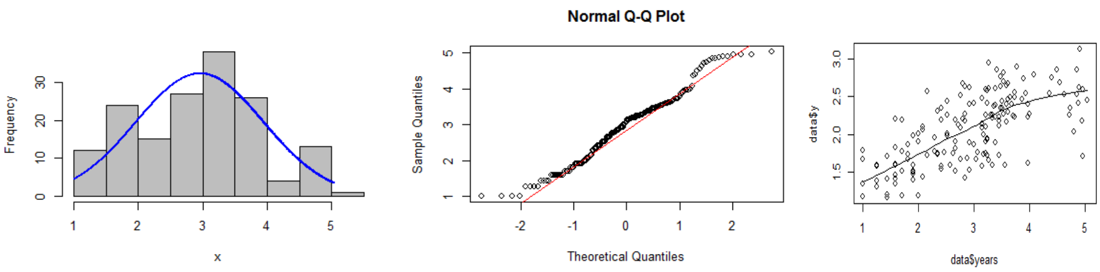
Tukey's Ladder of Powers transformation of the dependent variable Ridership



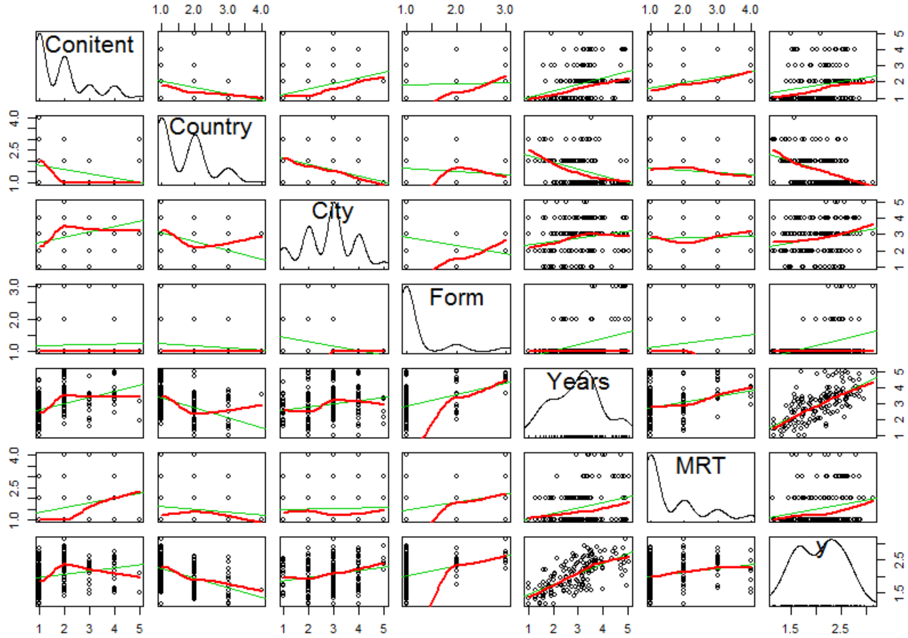
The distribution of the independent variable Years



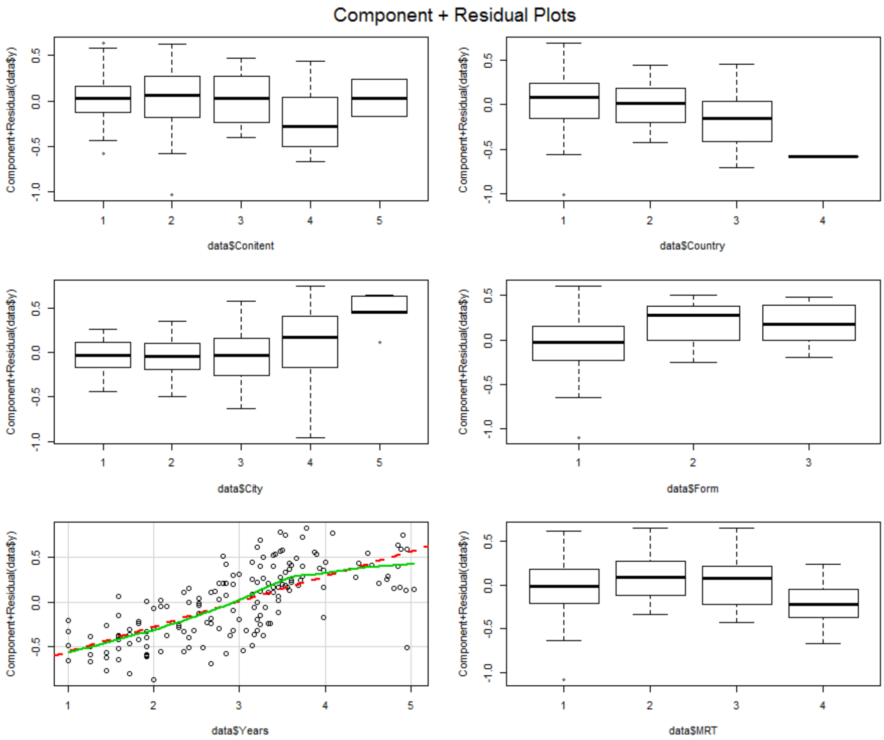
Cube root transformation of the independent variable Years



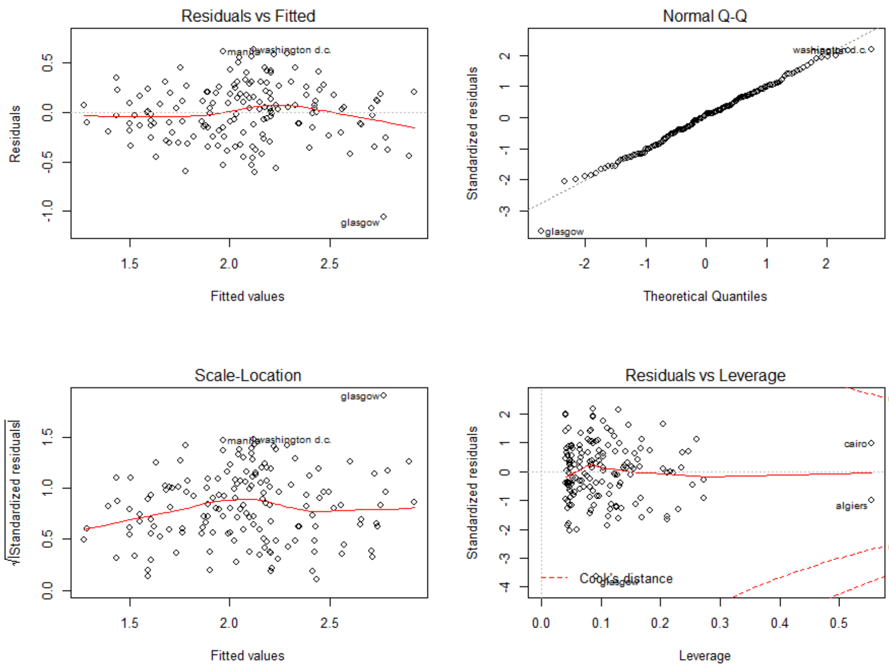
Model 1 Correlation Plots



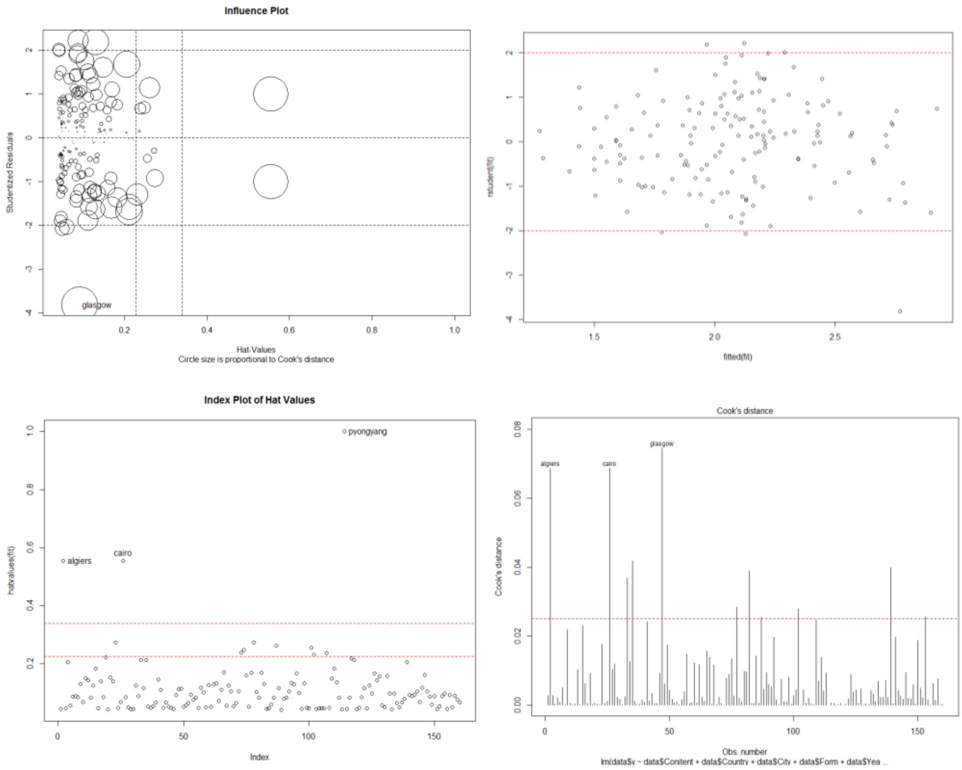
Model 1 Linear Relationship Plots



Model 1 Plots:

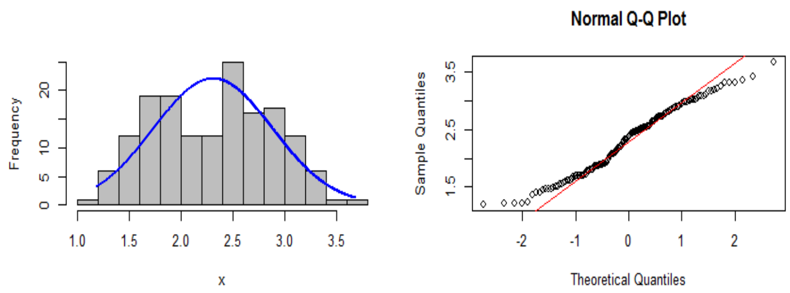


Model 1 Influence Plots:

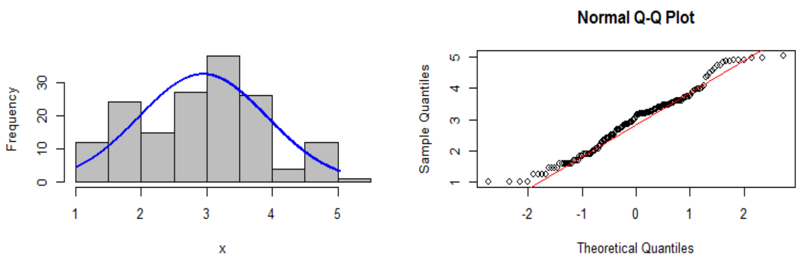


Final Model - Model 2 (Delete the point Glasgow)

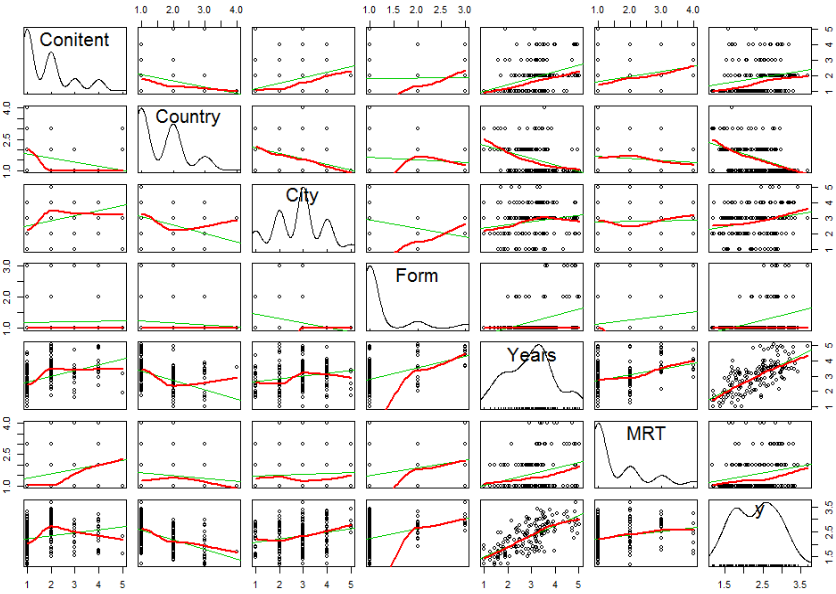
Tukey's Ladder of Powers transformation of the dependent variable Ridership



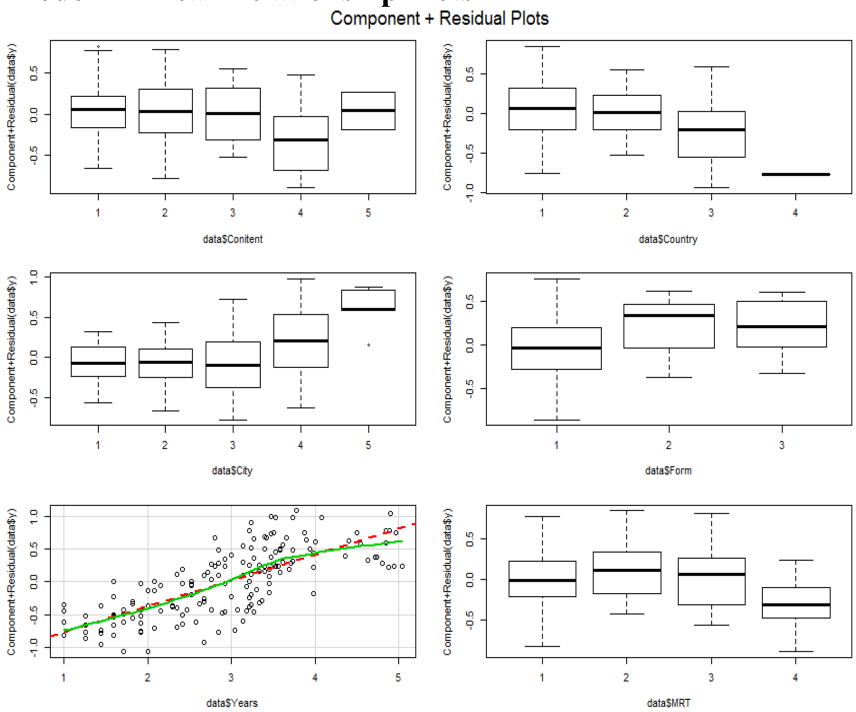
Cube root transformation of the independent variable Years



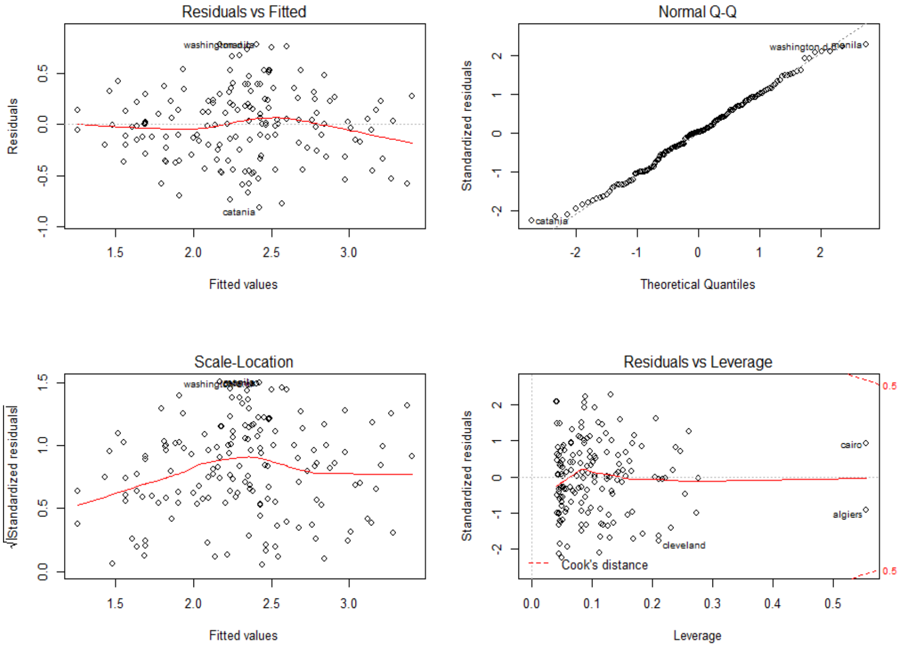
Model 2 Correlation Plots



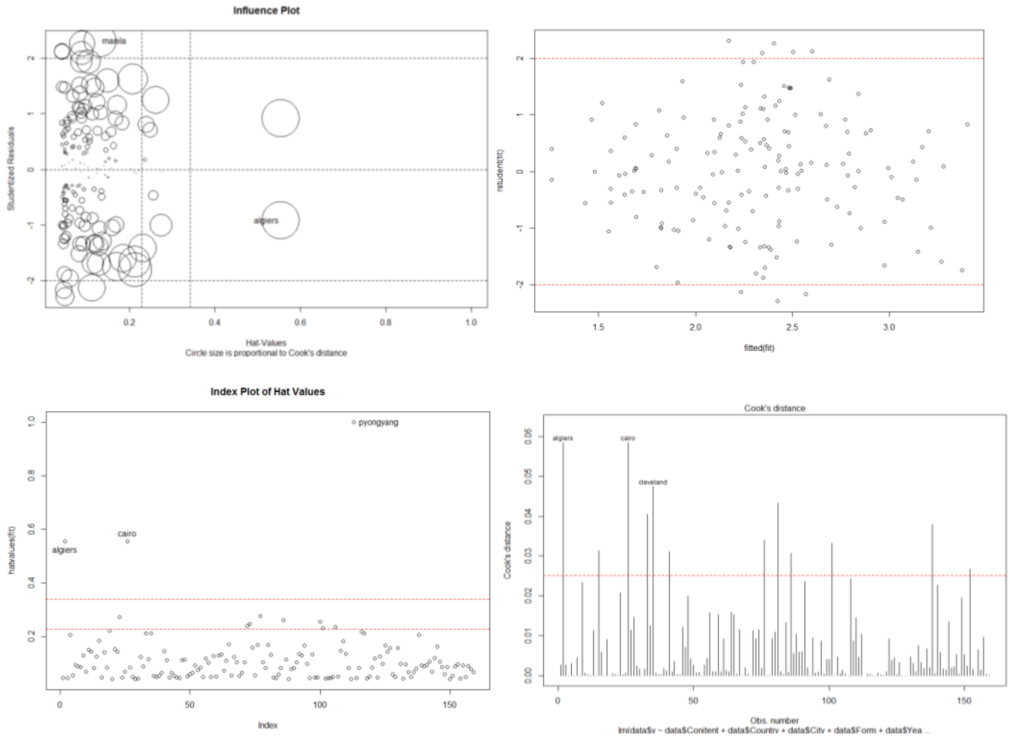
Model 2 Linear Relationship Plots



Model 2 Plots:



Model 2 Influence Plots:



Tukey's Ladder of Powers Transformation Results

	lambda	p-value
Model 1	0.175	0.006348
Model 2	0.2	0.00675

NCV Test Results

(Null hypothesis of independent of errors)

	Chi-square	Df	p-value
Model 1	4.416617	1	0.031664
Model 2	1.282816	1	0.257376

Durbin Watson Test Results

(Null hypothesis of independent of errors)

	Lag	Autocorrelation	D-W Statistic	p-value
Model 1	1	0.1039131	1.785014	0.148
Model 2	1	0.05946505	1.87367	0.408

VIF Test Results

Model 1

	GVIF	Df	GVIF^(1/2*Df)
Continent	4.632919	4	1.211245
Country	2.649654	3	1.176336
City	3.762398	4	1.180139
MRT	1.909899	3	1.113871
Form	1.992130	2	1.188036
Years	2.481908	1	1.575407

Model 2

	GVIF	Df	GVIF^(1/2*Df)
Continent	4.639645	4	1.211464
Country	2.646773	3	1.176123
City	3.801875	2	1.181680
MRT	1.925299	3	1.115363
Form	2.002868	2	1.189633
Years	2.545472	1	1.595453

Influential Observations

Model 1

	Studentized residuals	Hat	Cook's Distance
Glasgow	-3185268	0.09173778	0.07456164
Pyongyang	Na	1	Na

Model 2

	Studentized residuals	Hat	Cook's Distance
Algiers	-0.9190027	0.5544885	0.05846199
Manila	2.3129438	0.1305228	0.0432036
Pyongyang	Na	1	Na

Outlier Test Results

Model 1

	Studentized residuals	Unadjusted p-value	Bonferonni p-value
Glasgow	-3.815268	0.00020287	0.032257

Model 2

No Studentized residuals with Bonferonni p-value < 0.05.

References

AchARyA, S.R. and MoRichi, S., 2007. Motorization and role of mass rapid transit in East Asian megacities. IATSS research, 31(2), pp.6-16.

Allport, R.J., 1981. The costing of bus, light rail transit and metro public transport systems. Traffic Engineering & Control, 22(HS-032 837).

Angeloudis, P. and Fisk, D., 2006. Large subway systems as complex networks. Physica A: Statistical Mechanics and its Applications, 367, pp.553-558.

Ardila, A., 2007. How public transportation's past is haunting its future in Bogota, Colombia. Transportation Research Record: Journal of the Transportation Research Board, (2038), pp.9-15.

Barberillo, J. and Saldana, J., 2011. Navigation in large subway networks: An informational approach. Physica A: Statistical Mechanics and its Applications, 390(2), pp.374-386.

Barter, P., 1999. An international comparative perspective on urban transport and urban form in Pacific Asia: the challenge of rapid motorisation in dense cities (Doctoral dissertation, Murdoch University).

Ben-Akiva, M. and Morikawa, T., 2002. Comparing ridership attraction of rail and bus. Transport Policy, 9(2), pp.107-116.

Bertolini, L., 1996. Nodes and places: complexities of railway station redevelopment. European Planning Studies, 4(3), pp.331-345.

Bhandari, K., Kato, H. and Hayashi, Y., 2009. Economic and equity evaluation of Delhi metro. International Journal of Urban Sciences, 13(2), pp.187-203.

Boccaletti, S., Latora, V., Moreno, Y., Chavez, M. and Hwang, D.U., 2006. Complex networks: Structure and dynamics. Physics reports, 424(4-5), pp.175-308.

Cervero, R., 1994. Transit-based housing in California: evidence on ridership impacts. Transport Policy, 1(3), pp.174-183.

Cervero, R. and Murakami, J., 2009. Rail and property development in Hong Kong: Experiences and extensions. Urban Studies, 46(10), pp.2019-2043.

Chalermpong, S., 2007. Rail transit and residential land use in developing countries: hedonic study of residential property prices in Bangkok, Thailand. Transportation Research Record, 2038(1), pp.111-119.

Chan, A.P., Chan, D.W., Fan, L.C., Lam, P.T. and Yeung, J.F., 2008. Achieving partnering success through an incentive agreement: lessons learned from an underground railway extension project in Hong Kong. Journal of Management in Engineering, 24(3), pp.128-137.

Chan, S. and Miranda-Moreno, L., 2013. A station-level ridership model for the metro network in Montreal, Quebec. Canadian Journal of Civil Engineering, 40(3), pp.254-262.

Cheung, E., Chan, A.P., Lam, P.T., Chan, D.W. and Ke, Y., 2012. A comparative study of critical success factors for public private partnerships (PPP) between Mainland China and the Hong Kong Special Administrative Region. Facilities, 30(13/14), pp.647-666.

Cohen, B., 2006. Urbanization in developing countries: Current trends, future projections, and key challenges for sustainability. Technology in society, 28(1-2), pp.63-80.

Cornette, L., 2004. Contemporary neonatal transport: problems and solutions. Archives of Disease in Childhood-Fetal and Neonatal Edition, 89(3), pp.F212-F214.

Crotte, A., Graham, D.J. and Noland, R.B., 2011. The role of metro fares, income, metro quality of service and fuel prices for sustainable transportation in Mexico City. International journal of sustainable transportation, 5(1), pp.1-24.

Dablanc, L., 2007. Goods transport in large European cities: Difficult to organize, difficult to modernize. Transportation Research Part A: Policy and Practice, 41(3), pp.280-285.

Daganzo, C.F., 2010. Structure of competitive transit networks. Transportation Research Part B: Methodological, 44(4), pp.434-446.

Demery, Jr, L.W., 1994. Supply-side analysis and verification of ridership forecasts for mass transit capital projects. Journal of the American Planning Association, 60(3), pp.355-371.

Derrible, S. and Kennedy, C., 2009. Network analysis of world subway systems using updated graph theory. Transportation Research Record, 2112(1), pp.17-25.

Derrible, S. and Kennedy, C., 2010. Characterizing metro networks: state, form, and structure. Transportation, 37(2), pp.275-297.

Derrible, S. and Kennedy, C., 2010. Evaluating, Comparing, and Improving Metro

Networks: Application to Plans for Toronto, Canada. Transportation Research Record, 2146(1), pp.43-51.

Derrible, S. and Kennedy, C., 2010. The complexity and robustness of metro networks. Physica A: Statistical Mechanics and its Applications, 389(17), pp.3678-3691.

De Vos, J. and Witlox, F., 2013. Transportation policy as spatial planning tool; reducing urban sprawl by increasing travel costs and clustering infrastructure and public transportation. Journal of transport geography, 33, pp.117-125.

Dill, J., Schlossberg, M., Ma, L. and Meyer, C., 2013, January. Predicting transit ridership at the stop level: The role of service and urban form. In 92nd annual meeting of the Transportation Research Board, Washington, DC.

Duncan, M., 2010. To park or to develop: trade-off in rail transit passenger demand. Journal of Planning Education and Research, 30(2), pp.162-181.

Durning, M. and Townsend, C., 2015. Direct Ridership Model of Rail Rapid Transit Systems in Canada. Transportation Research Record: Journal of the Transportation Research Board, (2537), pp.96-102.

Guerra, E. and Cervero, R., 2011. Cost of a ride: the effects of densities on fixed-guideway transit ridership and costs. Journal of the american planning association, 77(3), pp.267-290.

Eriksson, L., Garvill, J. and Nordlund, A.M., 2008. Acceptability of single and combined transport policy measures: The importance of environmental and policy specific beliefs. Transportation Research Part A: Policy and Practice, 42(8), pp.1117-1128.

Flyvbjerg, B., Bruzelius, N. and van Wee, B., 2013. Comparison of capital costs per route-kilometre in urban rail. arXiv preprint arXiv:1303.6569.

Freilich, R.H. and Chinn, S.P., 1986. Transportation Corridors: Shaping and Financing Urbanization through Integratoin of Eminent Domain, Zoning and Growth Management Techniques. UMKC L. Rev., 55, p.153.

Fujii, S., Gärling, T. and Kitamura, R., 2001. Changes in drivers' perceptions and use of public transport during a freeway closure: Effects of temporary structural change on cooperation in a real-life social dilemma. Environment and Behavior, 33(6), pp.796-808.

Garrett, M. and Taylor, B., 1999. Reconsidering social equity in public transit. Berkeley Planning Journal, 13(1).

Goldman, T. and Gorham, R., 2006. Sustainable urban transport: Four innovative directions. Technology in society, 28(1-2), pp.261-273.

Gomez-Ibanez, J.A., 1985. A dark side to light rail? The experience of three new transit systems. Journal of the American Planning Association, 51(3), pp.337-351.

Gomez-Ibanez, J.A., 1996. Big-city transit rider snip, deficits, and politics: Avoiding reality in Boston. Journal of the American Planning Association, 62(1), pp.30-50.

González-Gil, A., Palacin, R. and Batty, P., 2013. Sustainable urban rail systems: Strategies and technologies for optimal management of regenerative braking energy. Energy conversion and management, 75, pp.374-388.

Gronau, W. and Kagermeier, A., 2007. Key factors for successful leisure and tourism public transport provision. Journal of Transport Geography, 15(2), pp.127-135.

Gilbert, R., Irwin, N., Hollingworth, B. and Blais, P., 2003. Sustainable Transportation Performance Indicators (STPI). Transportation Research Board (TRB), CD ROM.

Guerra, E., 2011. Valuing rail transit: comparing capital and operating costs with consumer benefits. Transportation Research Record, 2219(1), pp.50-58.

Gwilliam, K.M., 2002. Cities on the move: a World Bank urban transport strategy review. The World Bank.

Haghshenas, H. and Vaziri, M., 2012. Urban sustainable transportation indicators for global comparison. Ecological Indicators, 15(1), pp.115-121.

Hess, D.B. and Almeida, T.M., 2007. Impact of proximity to light rail rapid transit on station-area property values in Buffalo, New York. Urban studies, 44(5-6), pp.1041-1068.

Hine, J. and Scott, J., 2000. Seamless, accessible travel: users' views of the public transport journey and interchange. Transport policy, 7(3), pp.217-226.

Israel, E. and Cohen-Blankshtain, G., 2010. Testing the decentralization effects of rail systems: Empirical findings from Israel. Transportation Research Part A: Policy and Practice, 44(7), pp.523-536.

Jabareen, Y.R., 2006. Sustainable urban forms: Their typologies, models, and concepts. Journal of planning education and research, 26(1), pp.38-52.

Jun, M.J., Choi, K., Jeong, J.E., Kwon, K.H. and Kim, H.J., 2015. Land use characteristics of subway catchment areas and their influence on subway ridership in Seoul. Journal of Transport Geography, 48, pp.30-40.

Kain, J.F. and Liu, Z., 1999. Secrets of success: assessing the large increases in transit ridership achieved by Houston and San Diego transit providers. Transportation Research Part A: Policy and Practice, 33(7), pp.601-624.

Kamruzzaman, M., Baker, D., Washington, S. and Turrell, G., 2013. Residential dissonance and mode choice. Journal of transport geography, 33, pp.12-28.
Taylor, B.D. and Fink, C.N., 2003. The factors influencing transit ridership: A review and analysis of the ridership literature.

Kasarda, J.D. and Crenshaw, E.M., 1991. Third world urbanization: Dimensions, theories, and determinants. Annual Review of Sociology, 17(1), pp.467-501.

Kohn, H.M., 1999. Factors affecting urban transit ridership. Statistics Canada.

Lee, K., Jung, W.S., Park, J.S. and Choi, M.Y., 2008. Statistical analysis of the Metropolitan Seoul Subway System: Network structure and passenger flows. Physica A: Statistical Mechanics and its Applications, 387(24), pp.6231-6234.

Lee, S., Yi, C. and Hong, S.P., 2013. Urban structural hierarchy and the relationship between the ridership of the Seoul Metropolitan Subway and the land-use pattern of the station areas. Cities, 35, pp.69-77.

Levinson, D., 2012. Network structure and city size. PloS one, 7(1), p.e29721.

Litman, T., 2017. Introduction to multi-modal transportation planning. Victoria Transport Policy Institute.

Levinson, H., 2000. Rail transit in the next millennium: some global perspectives. Transportation Research Record: Journal of the Transportation Research Board, (1704), pp.3-9.

Li, X. and Lo, H.K., 2014. An energy-efficient scheduling and speed control approach for metro rail operations. Transportation Research Part B: Methodological, 64, pp.73-89.

Grava, S., 2003. Urban transportation systems. Choices for communities.

Lin, J.J. and Shin, T.Y., 2008. Does transit-oriented development affect metro ridership?: Evidence from Taipei, Taiwan. Transportation Research Record: Journal of the Transportation Research Board, (2063), pp.149-158.

Loo, B.P., Chen, C. and Chan, E.T., 2010. Rail-based transit-oriented development: lessons from New York City and Hong Kong. Landscape and Urban Planning, 97(3), pp.202-212.

Mackett, R.L. and Edwards, M., 1998. The impact of new urban public transport systems: will the expectations be met?. Transportation Research Part A: Policy and Practice, 32(4), pp.231-245.

Meyer, J.R., Kain, J.F. and Wohl, M., 1966. urban transportation problem.

Mitric, S., 1997. Approaching metros as potential development projects.

Mohammadi, A., Amador-Jimenez, L. and Nasiri, F., 2018. Review of asset management for metro systems: challenges and opportunities. Transport Reviews, pp.1-18.

Nieuwenhuijsen, M.J., Gomez-Perales, J.E. and Colvile, R.N., 2007. Levels of particulate air pollution, its elemental composition, determinants and health effects in metro systems. Atmospheric Environment, 41(37), pp.7995-8006.

Ovenden, M., 2015. Transit maps of the world. Penguin.

Pagliara, F. and Papa, E., 2011. Urban rail systems investments: an analysis of the impacts on property values and residents' location. Journal of Transport Geography, 19(2), pp.200-211.

Pickrell, D.H., 1989. Urban rail transit projects: forecast versus actual ridership and costs. Final Report (No. UMTA-MA-08-9021-89-1). United States. Urban Mass Transportation Administration.

Pojani, D. and Stead, D., 2015. Sustainable urban transport in the developing world: beyond megacities. Sustainability, 7(6), pp.7784-7805.

Pucher, J., 1988. Urban public transport subsidies in Western Europe and North America. Transportation Quarterly, 42(3).

Pucher, J. and Lefèvre, C., 1996. The urban transport crisis in Europe and North America. Springer.

Pucher, J. and Kurth, S., 1995. Verkehrsverbund: the success of regional public transport in Germany, Austria and Switzerland. Transport policy, 2(4), pp.279-291.

Richardson, B.C., 2005. Sustainable transport: analysis frameworks. Journal of transport geography, 13(1), pp.29-39.

Rissel, C., Curac, N., Greenaway, M. and Bauman, A., 2012. Physical activity associated with public transport use—a review and modelling of potential benefits. International journal of environmental research and public health, 9(7), pp.2454-2478.

Roth, C., Kang, S.M., Batty, M. and Barthelemy, M., 2012. A long-time limit for world subway networks. Journal of The Royal Society Interface, p.rsif20120259.

Siemiatycki, M., 2006. Message in a metro: building urban rail infrastructure and image in Delhi, India. International Journal of Urban and Regional Research, 30(2), pp.277-292.

Sinfield, J.V. and Einstein, H.H., 1998. Tunnel construction costs for tube transportation systems. Journal of construction engineering and management, 124(1), pp.48-57.

Sohn, K. and Shim, H., 2010. Factors generating boardings at metro stations in the Seoul metropolitan area. Cities, 27(5), pp.358-368.

Taylor, B.D., Miller, D., Iseki, H. and Fink, C., 2003. Analyzing the determinants of transit ridership using a two-stage least squares regression on a national sample of urbanized areas.

Taylor, B.D., Miller, D., Iseki, H. and Fink, C., 2009. Nature and/or nurture? Analyzing the determinants of transit ridership across US urbanized areas. Transportation Research Part A: Policy and Practice, 43(1), pp.60-77.

Tian, Z., Weston, P., Zhao, N., Hillmansen, S., Roberts, C. and Chen, L., 2017. System energy optimisation strategies for metros with regeneration. Transportation Research Part C: Emerging Technologies, 75, pp.120-135.

Topp, H. and Pharoah, T., 1994. Car-free city centres. Transportation, 21(3), pp.231-247.

Vuchic, V.R. and Musso, A., 1991. Theory and practice of metro network design. Public transport international, 3, p.298.

Wright, L. and Fjellstrom, K., 2003. Mass transit options.

Yang, X., Li, X., Gao, Z., Wang, H. and Tang, T., 2013. A cooperative scheduling model for timetable optimization in subway systems. IEEE Transactions on Intelligent Transportation Systems, 14(1), pp.438-447.

Zhan, X., Tan, Z.A., Domercq, B., An, Z., Zhang, X., Barlow, S., Li, Y., Zhu, D., Kippelen, B. and Marder, S.R., 2007. A high-mobility electron-transport polymer with broad absorption and its use in field-effect transistors and all-polymer solar cells. Journal of the American Chemical Society, 129(23), pp.7246-7247.

Zhang, J., Xu, X., Hong, L., Wang, S. and Fei, Q., 2011. Networked analysis of the Shanghai subway network, in China. Physica A: Statistical Mechanics and its Applications, 390(23-24), pp.4562-4570.

Zhao, J., Deng, W., Song, Y. and Zhu, Y., 2013. What influences Metro station ridership in China? Insights from Nanjing. Cities, 35, pp.114-124.