Radboud Universiteit Nijmegen

Nijmegen School of Management

Master Thesis Human Geography: Economic Geography

Externality Interdependencies: An Experimental Framework on the Economic Integration of Path Dependency on Externalities

Applying Multigroup Analysis in Structural Equation Modelling and the case of the Greater Pearl River Delta

By L.B.I. Lemmens (s4422260)

Internship: *Consulate General of the Netherlands in Hong Kong & Macao SAR*
Supervisor internship: *Rogier Hekking*

Supervisor Radboud University: *Dr. Martin van der Velde*

11-11-2020

Acknowledgements

I am very grateful to all people who helped and supported me in carrying out this project. The completion of this thesis would not have been possible without them.

First and foremost, I would like to thank all my former colleagues at the consulate of the Netherlands in Hong Kong for their kindness, support, assistance and enjoyable conversations during my entire internship. With a special thanks to Rogier Hekking and Annemieke Ruigrok, for providing me with the opportunity to get an insight into the consulate and involving me in a variety of activities and responsibilities at the consulate and beyond. I especially appreciate that they took the effort to let me finish my internship in Hong Kong after the outbreak of Covid-19.

Also, I would like to thank Martin van der Velde for his guidance and feedback that really enlightened this research. Due to his supervision, my research turned way more realistic and comprehensible, and mostly, more interesting.

And last but not the least, I am thankful to my parents and my friends for helping me finish my thesis in, already, quite a stressful year. Especially to Dennis, Kees and Lieve for all their motivational speeches during any of our daily coffee breaks at the university library - even though they were mostly sarcastic.

Table of contents

Acknowledgements II

Abstract IV

List of abbreviations Fout! Bladwijzer niet gedefinieerd.

List of tables and figures VI

Summary VII

1. Introduction 1

2. Theoretical framework 4
2.1 Basis of agglomeration economies 4
2.2 Externalities 5
2.3 Marshall versus Jacobs? 8
2.4 Path dependency and economic interaction 10
2.5 Externality interdependencies 12
2.5.1 Endogenous economic growth 14
2.5.2 Institutional context 15
2.5.3 Geographical proximity versus spatial connectivity 17
2.5.4 Economic industry structure 18
2.6 Econometric model 19

3. Case study: Greater Pearl River Delta 21
3.1 Regional and historical developments 22
3.2 Mega-city agglomerations in the world 27

4. Methodology and operationalization 30
4.1 Data collection 30
4.1.1. Spatial connectivity 32
4.1.2. Institutional distance 36
4.1.3. Industry similarity 41
4.2. SEM: externality interdependencies framework 43
4.3. Descriptive statistics and data analysis 47

5. Analysis and results 53
5.1. General model specifications 53
5.2. Multigroup analysis 54

6. Conclusion and discussion 59

7. References 62

Appendices Fout! Bladwijzer niet gedefinieerd.
A. Inter-city railway system and expressways in GPRD. 71
B. Annual improvements of travel times between regions in GPRD. 72
C. Industry similarity specifications. 73
D. Normality distribution of residuals. 74
E. Correlation coefficients of observed variables for each latent structure. 75
F. Normality distribution of residuals. 76

Abstract

The aim of this study is, first, to extend standard agglomeration and endogenous growth methodology by highlighting the path dependent perspective of externalities, and second, by developing an experimental multivariate framework for empirically measuring these externalities. The importance of path dependency in agglomeration studies is rarely discussed, while it is a main determinant for bilateral externalities. To understand the dynamic relations surrounding these concepts, we introduce the notion of 'externality interdependencies'. Where 'externality' refers to the complementary role of externalities in general, 'interdependencies' refers to the path dependent aspect of economic integration. This allows us to set up an econometric and statistical framework to investigate the dynamic relations of institutional distance, geographical proximity and industry similarity on 'externality interdependencies'. Using data of eleven prefecture-level cities within the Greater Pearl River Delta agglomeration, an experimental framework of a multigroup SEM capturing city-specific bilateral 'externality interdependencies' is developed. This MSEM allows testing if externalities have significant differences based on their city-specific parameters, as well as effects between pre-specified concepts. Unfortunately, estimations and results for confirming the proposed dynamic hypotheses are lacking due to empirical under-identification. Several resolutions and revisions are proposed. Nevertheless, the outlined SEM framework is practicable for several applications regarding agglomeration studies, as well as the empirical understanding of measuring regional externalities.

Keywords

Dynamic externalities; path dependency; externality interdependencies, structural equation modelling; multigroup analysis; Greater Pearl River Delta

List of abbreviations

AK	Accumulation of Knowledge
CCP	Chinese Communist Party
CFA	Confirmatory Factor Analysis
CNY	Chinese Yuan (¥)
FTZ	Free Trade Zone
GPRD	Greater Pearl River Delta
- Do	Dongguan
- Fo	Foshan
- Gu	Guangzhou
- Hu	Huizhou
- Ji	Jiangmen
- MSAR	Macau Special Administrative Region
- HKSAR	Hong Kong Special Administrative Region
- Sh	Shenzhen
- Zha	Zhaoqing
- Zho	Zhongshan
- Zhu	Zhuhai
GSB	Guangdong Statistical Bureau
HZMB	Hong Kong-Zhuhai-Macau Bridge
MAR	Marshall-Arrow-Romer
ML	Maximum Likelihood
MSEM	Multigroup Structural Equation Model
OCTS	One Country, Two Systems
PRC	People’s Republic of China
PRD	Pearl River Delta
SAR	Special Administrative Region
SEZ	Special Economic Zone
SEM	Structural Equation Model
SSEZ	Shenzhen Special Economic Zone

List of tables and figures

Tables

Table 3.1 - Major economic indicators of the Greater Pearl River Delta (2018). ***23***

Table 3.2 - Greater Pearl River Delta compared to other global mega-city agglomerations**.** ***29***

Table 4.1 - Average intercity travel times in 2000, 2010 and 2020 between Eastern, Central and Western regions in GPRD ***33***

Table 4.2 - Matrix of travel times between cities in the GPRD. ***34***

Table 4.3 - Measurement items of legal institutions, cultural institutions and normative institutions ***38***

Table 4.4 - Composite indexes for legal, normative and cultural institutional distance for Guangdong, HKSAR and MSAR. ***41***

Table 4.5 - Values for industry disparities for 'manufacturing' and 'tertiary' for 2010. ***43***

Table 4.6 - Conditions of all variables for structural equation modelling. ***45***

Table 4.8 - Pearson's correlation coefficients matrix of exogenous, observed variables. ***50***

Table 4.7 - Descriptions, descriptive statistics and values of the observed variables of interest for externality interdependencies ***52***

Table 5.1 - Description of multigroup baseline models considered. ***55***

Table 5.2 - Alternative MSEM group constraints with corresponding sample sizes in parenthesis. ***58***

Figures

Fig. 2.1 - Path dependent economic interaction of macro-, meso- and micro-level industry and agglomeration economies . ***12***

Fig. 2.2 - Theoretic framework for externality interdependencies. ***13***

Fig. 3.1 - Map of Greater Pearl River Delta, HKSAR and MSAR. ***22***

Fig. 3.2 - Gross Domestic Product of the Greater Pearl River Delta from 2000 till 2018. ***27***

Fig. 4.1 - Path diagram of the structural equation model for externality interdependencies. ***46***

Fig. 5.1 - Path diagram of the estimated general structural equation model for externality interdependenc. ***53***

Summary

Within economic and geographic studies, it is acknowledged that agglomerations in general have a significant positive dependency on local endogenous economic growth. This endogenous economic growth is generated by dynamic and interconnected spatial interactions, restricted to a certain geographical proximity which is referred to the agglomeration. These interactions are economically described as externalities, which can either be considered as negative or positive. Within agglomerations and related to endogenous economic growth, externalities often create incremental benefits due to the formation of spatially bounded comparative and competitive advantages (Dogaru et al., 2011). For instance, a homogeneous labour supply increases the total labour productivity, and a greater local interfirm trade surges knowledge creation and fosters innovations.

In general, externalities can be divided based on two types of geographically bounded firms and their related transmission channels: related and supporting industries. The first relates to the presence of homogenous and specialized companies, which states that endogenous economic growth and comparative advantages are created due to similar (or closely related) economic activities (Delgrado et al., 2016; Warf, 2010). These homogenous intra-industry externalities are often called Marshall-Arrow-Romer (MAR) externalities. The latter refers to the role of supporting industries on a firm's productivity, which generate differentiated knowledge spillovers (Illy et al., 2009). These externalities are often referred to as Jacob's externalities. This study does not contribute to the enduring debate which form of externality creates the highest endogenous economic growth, but views both externalities as mutually bounded (Jackson, 2015). Therefore, externalities are measured as both MAR and Jacob's externalities.

Nevertheless, the dynamic development of externalities needs to be highlighted. Within agglomeration studies, the dynamic role of externalities is often disregarded, even though it is the main determinant for economic integration. In order to include this economic interrelatedness, this study refers to the presence of path dependency. Path dependency states that regional historical processes, institutional contexts and local economic developments are the main determinants for local embeddedness (Djelic & Quacks, 2007; Martin & Sunley, 2006). Therefore, the existence and formation of externalities within an agglomeration is a dynamic process of several transmissions and channels between cities in an agglomeration, which are effected by historical sequences (Press, 2006). In order to refer to this dynamically formed economic interrelatedness, this study implements the concept of 'externality interdependencies'. This concepts captures both the complementary role of externalities in general ('externality'), as well as the path dependent aspects of economic integration ('interdependencies').

Based on externality interdependencies, the dynamic relations of endogenous economic growth within an agglomeration can be theoretically and empirically analysed. In order to analyse the specific externality interdependencies within an agglomeration, this study aims at deriving bilateral externality interdependencies between the cities of an agglomeration. This concept enables us to review the main research question of this thesis, which states 'do bilateral externality interdependencies effect endogenous economic growth within the Greater Pearl River Delta?'. In addition, this study includes three major determinants for externalities, namely institutional distance, spatial connectivity and industry similarity, to investigate the dynamic role of these specified variables in the dynamic setting of externalities.

Previous studies have analysed the theoretical relations of these concepts with regards to externalities in agglomerations, which provides a provisional base for the dynamic setting and empirical analysis. Externalities in general are presupposed to positively effect endogenous economic growth (Klenow & Rodriquez-Clare, 2005). It is theorized that increased economic interrelatedness creates

enhancing interdependencies as endogenous economic growth in agglomerations and economic size are directly related (Chen, 2002). Greater economic growth in an interrelated local environment fosters the presence of externalities, creating a supportive and cumulative circle of mutual connectedness. Besides the effect of economic growth in general, the effect of the previously mentioned concepts are theorized.

Firstly, similar institutional contexts are expected to improve the presence of externality interdependencies due to similar relational assets. Relational assets, which include norms, behaviour and cultural values, are present in each socio-cultural environment and are present in all layers of society. If relational assets between cities are similar it is expected to establish trust and reciprocity, which in line effects the ease of conducting business and a mutual sense of understanding and interactions (Amin, 1999; Storper, 1997). Therefore, a greater institutional distance between cities in an agglomeration is expected to negatively effect the presence of externality interdependencies.

Secondly, spatial connectivity is a widely acknowledged positive determinant for externalities in agglomeration (Burger & Meijers, 2016; Cappelo, 2000; Pessoa, 2011). Nevertheless, it is important to consider the dynamic role of spatial connectivity in relation with the other concepts in discussion. As elaborated by Gertler (2005), similar institutional contexts are often spatially bounded or developed which theorizes a complementary relationship between spatial connectivity and institutional distance on externality interdependencies.

Thirdly, the role of similar industry structures between cities in an agglomeration is discussed. As stated by Shihe, Xiaofend and Guojon (2010), similar industry structures firstly attract a specific interconnected economic environment. Consequently, this attractive economic environment promotes enhanced economic conditions for related and supporting industries and thus creating a diversified local economy. It is therefore hypothesized that (early) specialization has a significant and positive effect on the presence of externality interdependencies due to its promoted bilateral economic interconnectedness.

Based on this notion of externality interdependencies, the previously mentioned relationship and the theoretic framework, an endogenous simultaneous econometric model is formulated. Equations (2.4 and 2.5) provide a framework to measure externality interdependency and allow to empirically examine the role of various concepts in a dynamic setting. This study prefers to empirically analyse externality interdependencies from a bilateral perspective (see eq. 2.6a and 2.6b) and utilizing the Greater Pearl River Delta as agglomeration of study.

The Greater Pearl River Delta (GPRD) includes nine prefecture-level cities of the Guangdong province and two Special Administrative Regions: Macau (MSAR) and Hong Kong (HKSAR). The GPRD is one of the greatest economic agglomeration by means of GDP and population, and includes a strong diversified economy with (multiple) financial, political, cultural, transportation and economic centres in the agglomeration (Tang & Ellison, 2019). An extensive historical and developmental analysis is given which highlights the uniqueness of the global multi-industry economic hub.

This historical overview indicates the importance of path dependency within an agglomeration as it highlights that the GPRD is unquestionably divided, consisting of different regimes, economies, and industrial structures due to historical developments and interventions (Vogel et al., 2010). For instance, HKSAR and MSAR are two former colonies (of respectively Great Britain and Portugal) which results in both cities having different legislative structures, monetary systems and tax regulations. Of the nine prefecture-level cities in the Guangdong province, two cities are developed as competing 'Special Economic Zones' (SEZ) and three cities have international 'Free Trade Zones' (FTZ). These strongly conflicting but also mutualising historical developments are of great importance to understand the current (economic) interconnectedness in the agglomeration.

To statistically and empirically derive the bilateral externality interdependencies in the GPRD as proposed before, this study implements an experimental multigroup structural equation model (MSEM). This multivariate statistical method allow to measure the dynamic constructs of externality interdependencies, as it considers the essential mutual interconnectedness between the concepts of interests (Harlow, 2014). Within SEM concepts can be incorporated as latent variables, allowing to construct an unobserved variable from several manifest variables. Often within agglomeration studies, indicators for externalities are not directly measurable and need to be constructed from a subset of indicators. In addition, as this study focusses on bilateral externality interdependencies, a multigroup approach allows measuring city-specific externality interdependencies. This generates the possibility to perform group comparisons and analyse the specific role of cities in an agglomeration, besides plenty of alternative applications.

First, the previously mentioned concepts are operationalized in a bilateral setting for the GPRD between 2000 and 2018. Institutional distance is measured using as a combination of three measurable subsets of institutional distance, namely legislative, normative and cultural institutional distance (Bae & Salomon, 2010). Based on several indicators of the World Economic Forum (2019), The Heritage Foundation (2019) and Hofstede (2001), index values are generated for each city. Consequently, these three index values construct the latent variable for institutional distance. Regarding spatial connectivity, aggregate travel distances between each city in the GPRD are developed (Huo & Li, 2011). In order to eliminate biased data, major infrastructural developments in the agglomeration are analysed. Lastly, industry similarity is measured as industry deviation. Including the two most contemporary industries in the region (manufacturing and tertiary industry) allows to statistically generate differences between each city (Marll, 2008). For each industry category the percentage of GDP is taken and subtracted between the two cities of study, generating a measure for industrial deviation between both cities. In addition and based on previous empirical studies regarding externalities in agglomerations, several control variables are incorporated. These include population, wages, human capital, and FDI.

After analysing and assessing the data for statistical misconducts like multicollinearity, heteroskedasticity, normal distributions and outliers, the general SEM with 861 parameters and a degree of freedom of 767 is specified (see fig. 5.1). Identifying the SEM and applying Maximum-Likelihood (ML) estimations generate estimations. However, as the operationalization and methodology is constructed from a bilateral perspective, these results are not valid. Since no multigroup analysis is implemented, data is contradicting due to its bilateral structure. Therefore, a multigroup approach is required. Applying a configural (or equivalent forms) multigroup analysis in the structural equation model allows to test these single group cases and calculate factor loadings for each specific cluster (city in the GPRD) (Bou & Satorra, 2010).

Unfortunately, the configural MSEM is unable to estimate parameters in AMOS. A multistep analysis of possible estimation issues is conducted. Firstly, software issues are ruled out due to similar estimation issues in alternative statistical programs. This confirms that there is an improper solution for MSEM, caused by either nonconvergence, empirical underidentification, or un-identification (Kenny, 1979). As the SEM has generated feasible (but invalid) estimations, empirical underidentification is assumed to be the issue.

Since empirical underidentification in this study is caused by a non-positive sample matrix, several possible causations are investigated (Blunch, 2012). These include data entry errors, disturbing linearity or correlation coefficients, or issues regarding the sample size. Eliminating a subset of manifest variables in the measurement model to eliminate linearity or fixing several parameters between equivalent models do not generate feasible estimations. In addition, applying alternative estimations techniques like Bayesian estimations and unweighted least squares estimations do not generate feasible estimations. Therefore, it is assumed that the empirically minor multigroup sample size is not sufficient to create sufficient covariance information for MSEM. Several alternative clusters are tested as an alternative. However, no estimations are derived.

Conclusively, since the provisional and general SEM represented in fig. 5.1 has generated (conceptually invalid) estimations, the path diagram of the model is conceptually valid. However, since no MSEM estimations can be generated, proper conclusions are lacking. Since this is an issue that is caused directly from the data, an extensive discussion regarding the applied methodology and operationalization is necessary.

Several factors could contribute to empirical underidentification which could have been eliminated when considering a dissimilar research focus. Whereas this study has focussed on the bilateral perspective of externality interdependencies, it might have contributed to linearity amongst variable measures, serial correlation due to bilateral interconnectedness, or other statistical issues due to the bilateral perspective. In addition, the configural multigroup analysis could have been the cause for insufficient covariance information as a result of a lack of adequate observations. Whereas a multigroup baseline model could increase the sample size drastically, an alternative methodology needs to be applied. Nevertheless, despite the absence of valid MSEM estimations, the implications of the theoretical and econometric framework for externality interdependencies are extensive.

1. Introduction

The role of externalities within endogenous growth theories and agglomeration economies has received increased devotion in the field of economics throughout the last decades (Beaudry and Schiffauerova, 2009; Glaeser et al., 1992; Illy et al., 2009; van der Panne, 2004). Whereas the presence of externalities within agglomeration economies is a fundamental and deep-rooted belief, the characteristics of these economic effects remain unresolved. The debate focusses on two components of externalities, namely specialization or diversification. The prior refers to MAR-externalities, based on the theories of Marshall (1920), Arrow (1962) and Romer (1986). They state that externalities occur due to homogeneity of the local industry structure, creating a complementary and consensually improving specialized economy. Opposed to the notion of MAR externalities and specialization stands the presence of Jacobs' externalities. Jacobs (1969) states that externalities are a result of heterogeneity and a diversified economic structure. Differentiated externalities caused by diversification within a local environment foster innovation since it *'enables cross-fertilization of ideas …* [and] *offers local labour force a broader mix of skills to working with new technologies'* ('Jacobs' Externalities,'' 2017, p. 2). Where a vast amount of theoretical and empirical studies have contested whether specialization (MAR) or diversification (Jacobs) externalities are the main justifications for agglomeration economies, the issue remains unresolved (Di Clemente et al., 2014; Ejermo, 2005; Illy et al., 2009; Jackson, 2015; van der Panne, 2004; van der Panne & van Beers, 2006).

A more grounded and theoretically instituted belief is the harmonizing presence of both specialization and diversification externalities within an agglomeration as a result of mutually established regional economic relationships (Beaudry & Schiffauerova, 2009; Sunley, 2006). Several representations for this supportive relationship of both externalities have been used, including 'diversified specializations' and 'multiple specializations' (Hong & Xiao, 2016; Malizia & Feser, 1999). However, the core of this complementarity is mutual amongst these studies, namely the evolved economic interactions and local linkages within the agglomeration (McCann & Arcs, 2011). This perspective states that externalities are dynamically created and dependent on complex and historical structures within an agglomeration.

This study concentrates on this, often neglected, field within agglomeration economies: endogenous dynamic economic integration. To capture these endogenous and dynamic effects on externalities, a path dependent approach is applied. The theory of path dependency assumes that time and history are major determinants for current processes and embeddedness (Hassink, 2005; Henning, 2018). So, based on several dynamically formed relations, historical developments and institutional patterns, current economic equilibriums and structures are formed (Djelic & Quacks, 2007).

Constructed on path dependency theory and the supportive occurrence of MAR and Jacobs' externalities, a unique approach in the field of agglomeration economies is introduced and applied. Based on the path dependent approach, the role of various channels and dimensions within the agglomeration can be investigated. Where cities within an agglomeration often dynamically converge towards each other, the process of economic interrelatedness and integration is not solely based on past developments in the specific city. It includes the overall agglomeration and forms an interconnected framework dependent on multiple channels and dimensions (Press, 2006). Based on this dynamic economic integration, the presence of externalities between cities fluctuates over time and in a variety of industries and sectors simultaneously. This results in a dynamic and intertwined appearance of externalities within a variety of dimensions and externality channels between cities. Therefore, agglomeration externalities affect economic integration on agglomeration, macro- and micro-industry levels, and it creates a form of interdependency between each city (Feld & Boyd, 2006).

In order to capture this specified role of path dependent externalities within an agglomeration and to avoid misinterpretations, the concept of 'externality interdependencies' is introduced. This concept summarizes the role of externalities within agglomeration economies from a path dependent approach. Where 'externality' refers to the complementary role of externalities in general, 'interdependencies' is a result of path dependent economic integration (Feld & Boyd, 2006; Sunley 2006). Based on this novel concept of 'externality interdependencies', the dynamic, intertwined and path dependent role of internal factors on externalities within an agglomeration can be examined. Nevertheless, the estimation of externality interdependencies in agglomeration economies is not uncomplicated, since the practicalization of the theoretical concept includes an abundance of variables.

To estimate externality interdependencies, this study implements an experimental and multi-applicable multigroup structural equation model (MSEM). A SEM is the appropriate statistical method for analysing this non-static multivariate theoretical framework since it allows for comprehensive possibilities and implementations (Harlow, 2014). Adaptations of the theoretical framework, represented in a simultaneous equation model, can be included effortlessly in the multivariate framework of the SEM, while the influence of specific measures or unobservable externality channels can be implemented by the creation of latent variables and simultaneous effects between these variables (Harlow, 2014; Sunley 2006). In addition, SEMs are applicable with multigroup analysis, allowing to measure specific outgoing and incoming externality interdependencies within an agglomeration. Thus, the MSEM is constructed as an experimental framework for measuring bilateral externality interdependencies, allowing numerous (theoretical) modifications and conversions for future applications, while approaching agglomeration economies from a path dependent perspective.

As path dependency states that an agglomeration's externality interdependencies are highly dependent on historical developments, regional channels and place-specific dimensions, the underlying structure of an agglomeration needs to scrutinized comprehensively. The developments and economic growth over the past 19 years (2000-2018) of the mega-agglomeration 'Greater Pearl River Delta' (GPRD) are analysed in this study. The GPRD compromises nine major prefecture-level cities in the southern part of the People's Republic of China, and the Special Administrative Regions Hong Kong and Macau. The region is globally seen as one of the greatest and sophisticated agglomerations due to its high level of economic interaction and historical developments, including two former European colonies and several economic zones (Tang & Ellison, 2019), which makes it a sophisticated agglomeration from a path dependent perspective. After describing historical developments and institutional varieties between the eleven cities within the agglomeration, it is necessary to compare the development of the GPRD to developments of other global mega-city agglomerations to explicitly illustrate the path dependent developments for externality interdependencies.

Thus, the experimental structural equation model, based on previously investigated theoretical grounds within agglomeration studies and in the case of the Greater Pearl River Delta, provides a framework for estimating externalities from a path dependent perspective: externality interdependencies. Applying a multigroup analysis results in the possibility to measure bilateral externalities, examining the endogenous economic integration between pre-specified cities.

Existing studies have failed to capture this interrelated role of externalities. Within the field of economic geography, it is a unique and multi-applicable approach for agglomeration studies. Along with statements of Beaudry and Schiffauerova (2009), van der Panne (2004), and Illy et al. (2009), the dynamic role and intertwined role of various dimensions and channels on externalities needs to be examined further, as well as the underlying relations between these dimensions. Besides this desired inclusion of pre-analysed theoretical relations and examining relations between these channels, the SEM provides a structure for future implications in this stated scientific gap. The experimental and simultaneous framework endorsed in this paper enables measuring potential models, hypotheses and relationships in agglomeration-, core-periphery - and other relevant (geographic) studies.

In addition, the underlying structure of externality interdependencies within agglomeration can be of utmost importance for policy implementations, research studies, location assessments and other relevant institutions examining an agglomerations' structure. From a corporate perspective, the role of local legislative and social variations can be main determinants for companies to conduct business within an agglomeration, while the values of externality interdependencies can be indicators for interesting nearby and highly connected cities with core-cities within the agglomeration. From a governmental view, policies focussing on economic growth can be examined dependent on this framework, as well as the effect of possible policy interventions on the total streams of externality interdependencies due to the broad applicability of the multigroup structural equation model.

So, the framework in this study analyses the basic foundation of the experimental SEM and includes several theorized channels and dimensions. This paper makes these described contributions to the scientific field of economic geography, by addressing the following research questions. The main research question, referring to the overall role of externality interdependencies, is as follows:

> '*Do bilateral externality interdependencies effect endogenous economic growth within the Greater Pearl River Delta between 2000 and 2018?*'

Due to the path dependent perspective of this study, several dimensions can be investigated in the experimental empirical framework. Three related concepts within agglomeration economies are examined, namely institutional context, intercity spatial connectivity, and local industry structure. These measures have been analysed theoretically (and to some extend empirically) in agglomeration studies (Di Clemente et al. (2014); Gerritse & Arribas-Bel, 2018; Pike et al., 2015). However, these analyses are analysed independently and measured as a singular and static determinant for externalities. In order to derive a consistent measure for the role of these specified concepts, these concepts are theorized from a path dependent perspective and empirically analysed in a dynamic, multidimensional framework. Since externalities are dynamically formed, each concept has a variant role when being placed in an economically integrated framework (Sunley, 2006). This results in the followings research questions:

> '*What is the dynamic role of institutional distance on bilateral externality interdependencies in the GPRD?*';
>
> '*What impact does spatial connectivity between cities in the GPRD have on bilateral externality interdependencies?*';
>
> '*How did industrial similarity between cities of the GPRD effect bilateral externality interdependencies?*'; and
>
> '*How have institutional distance, spatial connectivity and industrial similarity, simultaneously affected the presence of externality interdependencies in the GPRD, and how do they relate to each other?*'.

Consequently, these concepts need to be theorized from a path dependent approach and quantitively covered in a statistical procedure. In this study, an experimental structural equation model is applied which is grounded on the theoretical and empirical patterns between these concepts on externality interdependencies. Since the notion of externality interdependencies is new in geographic studies, the following research question needs to be included::

> ‘*Can externality interdependencies be measured by applying a multigroup structural equation model?*’.

The remainder of this thesis proceeds as follows. The next section gives an overview of the theoretical framework, the implementation of path dependency in agglomeration economies, and the theorized relations. The third chapter is devoted to the Greater Pearl River Delta, followed by a broad analysis of the methodology and operationalization. Specifically, methods for collecting data are elaborated, the accurate path diagram of the SEM is developed, and the data is analysed. The following section describes the model specifications and multigroup analysis, as well as model improvements to generate estimates. The last chapter discusses further improvements of the model, and the theoretical and practical implications of this study.

2. Theoretical framework

The following chapter will elaborate on the theoretical framework of this study. Before all else, the basic and predominant assumptions in agglomeration economies are elaborated, followed by a detailed description of the externalities within agglomerations. Hereafter, a path dependent approach is gradually implemented into this study. This results in the construction of ‘externality interdependencies’, which allows testing several specified theories present in agglomeration theories. Subsequently, an econometric model is specified based on the notion of externality interdependencies.

2.1 Basis of agglomeration economies

Traditionally, economists and geographers have analysed the role of spatial structures on economic growth and welfare effects. Often, these fields combine theories related to spatiality, core-periphery theories, the location of industries, and agglomeration economies. The latter, agglomeration economies, has received a growing interest during the last centuries in the field of economic geography (Dogaru et al., 2011). Whereas the effects of agglomeration economies can be seen as positive or negative depending on the field of research, theories within economic geography are consistently stating a significant positive dependency between agglomeration economies and local economic growth (McCann & Van Oort, 2019). Economic geographers state that ‘*agglomeration forces leading towards a dynamic and self-enforcing process of increased agglomeration, and higher levels of welfare of the population in these agglomeration*’. In addition, these ‘*welfare effects are generated by a ‘love of variety’ by consumers and a ‘supply of varieties’ that increases with the economic size of a region*’ (Dogaru et al., 2011, p. 487).

According to these statements and in line with an abundance of theories, several assumptions within economic geography are prevalent. Firstly, agglomeration theories are based on the theory of endogenous growth (Krugman, 1991; Solow, 1956). They state that economic growth is the result of dynamic and interconnected spatial interactions, bounded by a certain sense of geographical proximity. Key of their assumption relates to ‘*the pervasive influence of some kind of increasing returns’* (Krugman, 1991, p. 5) as the binding factor between endogenous economic growth and agglomeration economies.

Secondly, these interactions are formed dynamically and incoherently. This relates to a crucial and vital distinction within agglomeration economies, namely defining an agglomeration as static or dynamic (Camagni et al., 2017). The former, a static agglomeration, refers to the superior efficiency of a larger local economy, generating lower productivity or transportation costs. Related to dynamic agglomerations, Chertow et al. (2008) state that '*Dynamic agglomeration economies evolve from knowledge creation and learning over longer periods of time, and may be associated with a history of interactions and cultivated long-term relationships'* (p. 1302).

This distinction needs to be carefully analysed. Firstly, economic growth within an agglomeration is related to a variety of factors which interact to a certain degree (McCann & Arcs, 2011). Endogenous economic growth cannot be determined by a solitary factor. In agglomerations, it is a result of a conglomerate effect between each city, determined by multiple, simultaneously occurring, variables. This is in line with Camagni et al. (2017), who state that a dynamic agglomeration '*indicates the possible drivers of efficiency increases for each city size, especially in terms of the capacity to change a city's internal characteristics which may act as structural constraints on its growth'* (p. 134). These internal characteristics and possible drivers of efficiency depend on this previously mentioned conglomerate effect within agglomeration economies. A dynamic approach would thus allow to effectively analyse these factors - decreasing the sense for generalization – and validly measure (interacting) variables to endogenous economic growth, including a historical perspective (Camagni et al., 2017).

In addition, where core cities are the main attractors of innovation, economies of scale, and economic growth, periphery cities profit from these developments. For instance, Gruber and Soci (2010) discuss such opportunities of *'developments for rural areas: as economies grow, immobile factors such as land might become higher valued, which in turn opens up the opportunity for new activities in those rural areas, such as tourism and recreation'* (p. 65). These developments are a result of historical and dynamic processes and are associated with long-term relationships. Thus, a dynamic approach is necessary to capture these multiple interactions within an agglomeration (Chertow et al., 2008).

Therefore, a dynamic approach allows focusing on the broad presence of sectors or industries which are historically created due to an abundance of local urban activity. The dynamic and augmenting process of agglomeration economies fosters local innovation and productivity, increasing endogenous economic growth through externalities.

2.2 Externalities

Since early studies in agglomeration economies, externalities have been seen as one of the most significant factors of endogenous economic growth (Neffke et al., 2011). Externalities in general are direct or indirect costs or benefits which are caused by other parties. Negative externalities often relate to environmental issues such as health issues in a neighbourhood caused by pollution of a nearby firm. Positive externalities are incremental benefits which are encouraged due to the actions of third parties, like educational skills or the effect of public infrastructure (besides their original proposed tasks). Within agglomeration and economic studies, externalities are often seen as benefits as a result of specific production or consumptions within the agglomeration. Externalities create a comparative and competitive advantage as firms and companies in distinctive regions do not profit from these externalities. For instance, firms decisively specialize their supply chain, labour force and capital. A homogeneous supply chain improves the internal resource allocation and network-relation; division of labour improves individual knowledge, skills and utility, and thus total productivity; and efficient internal capital allocation increases productivity and firm performance (McNeil & Moore, 2005; Porter, 1990; Smith, 1776/2010). As stated by Porter (1990), numerous nearby firms have a historical enhanced effect on the previously mentioned production factors and thus the productivity of nearby firms. Widely acknowledged are the complementary effects of industrial density namely *'to receive information spilling over from other firms, to reduce transport costs of interfirm trade, and to enhance the diversity of firms and local products available'* (Henderson, 1997, p. 75). Furthermore, it is generally acknowledged that endogenous intra- and inter-industry spillovers results in tacit knowledge spillovers and is consequently dependent on spatial proximity (Marshall, 1920/2009; van der Panne, 2004).

Insights provided by Marshall (1920), Arrow (1962), Jacobs (1969) and Romer (1986) are central in externality studies. As a result of their additions, externalities can be divided depending on their position and transmission channels. Porter (1990) distinguishes these externalities based on two types of geographically bounded firms: related and supporting industries.

Related industries: Marshallian externalities

In line with the role of related industries, the presence of homogeneous specialized companies fosters cost and productivity advantages by various related linkages (Delgrado, Porter & Stern, 2016; Porter, 1998). One of the earliest studies that contributed to the study of positive externalities amongst specialised but related industries is Marshall (1920/2009). Marshall stated three externality channels within related industries: input-output linkages, labour market pooling and knowledge spillovers. The first two externalities are components of Marshall's (1920) concept of 'Industrial District-argument', which relates to *'asset-sharing, such as the provision of specific goods and services by specialized suppliers and the creation of a local labour market pool sustained by a local concentration of production'* (van der Panne, 2004, p. 594). Homogeneous firms are involved in similar production processes, involving several forward and backward economic linkages. The enhanced homogeneous demand fosters economic opportunities and innovative capabilities to anticipate on the less-generalized necessities, and consequently advancing to a specialized, more skill full, labour supply (Delgrado et al., 2016; Warf, 2010). In addition, geographically bounded but homogeneous knowledge spillovers enhance the local stock of knowledge, increasing a firm's ability to innovate. As stated by Audretsch and Feldman (2004), *'firms exist exogenously and then engage in the pursuit of new economic knowledge as an input into the process of generating innovative activity'* (p. 2). These inputs are caused by the exchange of tacit knowledge between proximate firms. As tacit knowledge is undocumented, non-formal and verbally transferred, geographical closeness is a crucial determinant of the existence of this externality. Furthermore, as these knowledge spillovers occur mostly on diverse and unanticipated occasions these benefits are costless and are not only beneficial to the local firms.

Over time several researchers have contributed to these notions, further strengthening Marshall's beliefs (Glaeser et al., 1992). Arrow (1962) contributed some six decades ago by stating that geographical closeness for tacit knowledge spillovers to occur is of greater importance than previously assumed. Furthermore, he empirically found that tacit knowledge spillovers are non-rivalrous in practice. This ruled out criticisms that tacit knowledge spillovers do not occur between rivalling firms (Arrow, 1962; Audretsch & Feldman, 2004).

Romer (1986) supplemented to this concept by transforming Solow's (1956) economic growth model of accumulation of knowledge, by highlighting knowledge as a factor for endogenous growth. He firstly formulated and adapted the widely acknowledged model of accumulation of knowledge. This model *'of endogenous technological change in which long-run growth is driven primarily by the* *accumulation of knowledge by forward-looking, profit-maximizing agents'* (Romer, 1986, p.1003), has the following simplified form

$$Y_t = AK_t \qquad (2.1),$$

where

$$\dot{K} = sY_t - \delta K_t \qquad (2.2).$$

This AK-model, simplified by Jones (2019), includes an exogeneous productivity parameter (*A*), an endogenous knowledge (or defined as human capital in subsequent studies) parameter (*K*), and an exogeneous, constant investment rate (*s*), which is depreciated over time (δ). In addition, Y_t represents economic growth and productivity and Romer explicitly states that *K* is interpreted as technological innovativeness, derived from the accumulation of standardized knowledge.

Combining these two equations, there results

$$gY = \frac{\dot{Y}_t}{Y_t} = sA - \delta \qquad (2.3).$$

Equation (*2.3*) shows that regional economic growth (*gY*), is equal to the future change in outcome ($\frac{\dot{Y}_t}{Y_t}$), and is solely determined by the productivity and investment rate. So, an increase in the investment rate fosters economic output and the level of productivity. Combining (2.*1*) and (2.*3*), Romer (1986) concluded that an increase in local productivity (*A*) fosters the accumulation of local knowledge (*K*). This creates a bilateral relationship of increasing returns between the local accumulation of knowledge and productivity, resulting in a permanent economic growth (Jones, 2019; Romer, 1986). These contributions of Arrow (1962) and Romer's (1986) factor of endogenous knowledge, have attributed to geographical homogeneous intra-industry externalities being referred to as Marshall-Arrow-Romer (MAR) externalities (firstly introduced by Glaeser et al. (1992)).

Correspondingly, these MAR externalities channels are correlated and complementary to localization economies. This contributes to the belief that these channels establish mutual '*advantages that result from a spatial concentration of companies operating in the same industry or conducting similar types of activities. Companies in similar industries benefit from co-location due to the creation of a regional pool of specialised inputs'* (Pessoa, 2011, p. 9). Localization economies state that proximate intra-industry relations generate extensive homogeneous (undetermined) collaboration on the labour market, value chain's and knowledge creation, increasing simultaneously a firm's innovation capability, competitive advantage and internal productivity. Dicken (2015) describes these specialized economic environments as a specialized cluster, or specialization, which *'reflect the tendency for firms in the same, or closely related, economic activities to locate in the same places'* (p. 69).

This consequently stimulates the occurrence of tacit knowledge spillovers, reinforcing the specialization of the intra-industry's value chains. This is also insinuated in Romer's (1986) AK model, which indicates that also vice versa regional homogeneous productivity growth enhances the establishment of tacit knowledge spillovers. It shows that each of the previously elaborated channels of specialization, input-output linkages, labour market pooling and tacit knowledge spillovers, are complementary and support each other. In addition, a comprehensive collaboration structure of homogeneous firms allows the creation of specialised shared economies of scale, leading to downward sloping average cost curves and reducing capital entry barriers for new homogeneous (innovative) companies (Pessoa, 2011). Consequently, localization of an abundance of intra-industry firms results in specialized economic environments. Due to these complementarities between the MAR externalities occurring between homogeneous firms, this study refers to all MAR externalities cooperatively when discussing specialization.

Supporting industries: Jacobs' externalities

Besides related industries, Porter (1990) elaborates the role of supporting industries on a firm's productivity. The presence of economically diversified industries within a close proximity has firstly been elaborated by Jacobs (1969). Geographical supporting and complementary industries are the main determinants on a firm's level of innovation due to the creation of differentiated knowledge spillovers. Heterogeneous industries enhance these spillovers '*as diversity enables the cross-fertilization of ideas. In addition, diversity offers local labour force a broader mix of skills to working with new technologies'* (''Jacobs' Externalities,'' 2017, p. 2).

Jacobs' externalities are closely related to the concept of urbanization economies, which concentrates on the density of inter-industry linkages and the diversified pattern within the regional economy (Illy et al., 2009). It is important to acknowledge that the similarities and thus the relatedness between Jacobs' externalities and urbanization economies are debatable (Beaudry and Schiffauerova, 2009; Neffke et al., 2011). Urbanization economies mainly focus on an intensified urban economy of scale, providing numerous benefits to firms operating in the regional economy. However, urbanization economies also stimulate negative externalities leading to higher cost environments: expensive labour, higher land and rent prices, and more expensive basic living consumptions.

In this study, urbanization economies and diversification are relatable and supportive. As economies of scale lead to large-scale markets the overall costs for transportation and infrastructure decrease. Consequently, these economies of scale result in a greater industrial environment, enriching *'a more diversified industrial structure…* [and thus] *different and complementary technological knowledge and therefore favour innovative activities'* (Illy et al., 2009, p. 6). As the review study of Beaudry and Schiffauerova (2009) illustrates, the proposed negative externalities of urbanization economies are significantly marginalized compared to the positive externalities of economies of scale. In addition, urbanization economies promote the creation of Jacobs' externalities as firms benefit from greater urban diversity (Fu & Hong, 2011). This is in line with Dicken's (2015) concept of diversified clusters, which consist of large heterogeneous aggregate demand to reduce average economic costs and promote inter-industry knowledge spillovers and innovations.

2.3 Marshall versus Jacobs?

Undoubtedly, MAR and Jacobs' externalities must not be presumed to be mutually exclusive. It is improbable to assume that pure localization occurs in a specific region and pure urbanization in another region. Nevertheless, it is essential to theoretically analyse the impacts of MAR specialization and Jacobian diversification externalities and which are considered to be the most effective for regional economic growth.

A vast amount of theoretical and empirical studies have debated whether diversification or specialization externalities result in the most optimal dynamic economic growth (Di Clemente et al., 2014; Ejermo, 2005; Illy et al., 2009; Jackson, 2015; van der Panne, 2004; van der Panne & van Beers, 2006). Most studies resemble a significant positive effect of both described externalities on the regional economy, innovativeness or competitiveness. These are summarised by Beaudry and Schiffauerova (2009) in a review study. Their study shows that out of 67 studies, around 70% of the researches provided support for a positive relationship between MAR externalities, specialization, and a region's economic activities. In comparison, a similar 50 out of the 67 studies (75%) confirmed a positive correlation between Jacobs' externalities, diversification and regional economic activity. In addition, Jackson (2015) conducted a review study including more than 80 relevant papers leading to numerous conflicting conclusions about a specific preferred externality channel. Hence, these theoretical studies are ambiguous in indicating which type of externalities are determinedly present in regional economies.

Empirical studies also do not reach a consensus. Ejermo (2004) firstly provides a theoretical foundation in support of Jacobs' externalities. However, his empirical research indicated that only Marshallian externalities are present. An extensive empirical study on the effect of diversified and specialized economic structures in German cities by Illy et al. (2009) shows interesting results. Evidence shows that there is a U-shaped relationship between specialization and urban or regional economic growth, while diversification externalities appear to have no significant effect. In addition, MAR externalities only hold for cities and regions. States and counties are too immense to receive geographical knowledge spillovers based on associated knowledge. Furthermore, van der Panne and van Beers (2006) state that innovativeness is more promoted in specialized structures compared to diversified economies. However, Jacobs' externalities prove to have a long-term significant positive effect as firms in diversified regions outperform specialized firms after two years (van der Panne & van Beers, 2006).

As these studies show conflicting and ambiguous results based on various geographical units of observation, measures for centrality/diversity, and the research period, a preferred approach is unrecognized. Nevertheless, a similar consensus can be found amongst the previously described studies, disclosing a supportive relation between Jacobs diversification and MAR specialization externalities. Jackson (2015) concludes in his review study that *'I have identified a number of key dimensions of industry clusters and industrial diversification that suggest strongly that these two need not be strange bedfellows, but instead might well be very supportive partners'* (p. 123). This might also justify the absence of positive correlations in several studies summarized by Beaudry and Schiffauerova (2009). Considering that nearly all of these studies thus focus on specific but related externalities, enhanced corresponding effects might be ruled out and lead to insignificant results.

Several studies have theoretically investigated the connection between MAR and Jacobs' externalities, including Desrochers and Sautet (2008). They state that diversified regions are made up of various specialized regions. Malizia and Feser also shortly indicate this, stating that '*economic diversity is the presence of multiple specializations'* (1999, p. 92). In addition, it is mentioned by Guiliani (2003). She states that externalities due to centralized homogeneous firms are not only beneficial to the local firms, but also increase the potential and attractiveness of the entire region. Due to the creation of a virtuous and constant local economic growth, technological innovations and increased market conditions, the local economic environment becomes suitable for more miscellaneous industries and thus diversifies and intensifies (Guiliani, 2003).

Hong and Xiao (2016) formulate this supportive interaction between MAR and Jacobs' by using 'diversified specializations'. They state that their concept of 'diversified specializations' captures the relatedness between Jacobs'and MAR externalities. This is the case since a relatively small amount of minor homogeneous industries '*benefit from the economies of scale of each of the specialized sectors and at the same time enjoy the economies of scope and positive externalities from cross-fertilization across diverse sectors'* (Hong & Xiao, 2016, p. 4).

Based on these theories, specialised urban areas foster fertile ground for supporting industries in the long run. This generates the presence of Jacobs' externalities and diversified local economic activity. On the other hand, effective geographical diversification fosters regional innovation capabilities, reinforces industry competitiveness and develops an augmented homogeneous economy (Illy et al., 2009). Henceforth, the economic environment is stimulated by the increased presence of MAR externalities. Thus, it can be stated that this supportive connection between MAR and Jacobs' externalities results in a dynamic structure between specialization and diversification economies.

2.4 Path dependency and economic interaction

The dynamic structure and the importance of both sorts of externalities in relation to the development and establishment of economic growth is widely discussed in economic literature (Delgrado, 2016; Henderson, 1997; Porter, 1990). Braunerhjelm (2011) states that:

> *Growth can basically be attributed the following fundamental forces: an increase in factors of production; improvements in the efficiency of allocation across economic activities; knowledge; and the rate of innovation. Given full employment and efficient allocation, growth is thus driven by knowledge accumulation and innovation.* (p. 161)

In other words, endogenous regional economic growth is determined by the creation, collaboration and conglomerate effect of externalities, while assuming the absence of or controlling for exogeneous factors that interrupt full employment or efficient allocation. Furthermore, it is widely known that (most) entities in a regional economy are linked to each other and interact to a certain degree (McCann & Arcs, 2011).

Researchers have often tried to capture these economic interactions and dynamics. One approach to measure regional externalities is the input-output theory, which states (in a simple form) that the difference between local production and local consumption determines the regional interdependency for locally produced products (Wang, Dong & Mostafizi, 2017). However, this model does not include the spatial and network characteristics which are the main determinants of the current regional interlinkages (Zhang & Peeta, 2010).

Therefore, this study will relate to the theory of path dependency. According to Martin and Sunley (2006), path dependency is often disregarded in empirical studies, even though it is the main determinant and linkage for regional dynamic economic relationships. These relationships and externalities are determined due to several channels and dimensions. All of these macro- and micro level market effects, including geographical connectivity, horizontal and vertical linkages among homo- and heterogeneous industries, and institutional differences, affect MAR and Jacobs' externalities. So, on the macro level, the effect on regional economic growth is not dependent on a singular externality, yet on the sum of all these channels. In conclusion, externalities within a region are part of a complex structure due to this mixed set of internal distinctions (see *fig. 2.1*).

To understand a region's complex structure, both time and history are of great significance (Henning, 2018). Whereas geographical clusters are dynamically formed, the strengths and weaknesses of the institutional context and regional historical processes are determinants for the local embeddedness (Hassink, 2005). This is in line with Djelic and Quacks notion of path dependency, which '*characterizes historical sequences in which contingent events set institutional patterns with deterministic properties into motion'* (2007, p. 161). Its theory includes the assumption that the current economic equilibrium is a result of several dynamically formed relations on several dimensions.

It is important to mention the specific implementation of path dependency theory in this study as this theory has faced criticism and issues since its initial mentioning in geographic studies in the 1980s (David, 1985; Martin & Sunley, 2006). Firstly, researchers have regularly tried to analyse '*whether there are different types and degrees of path dependence ('historicity') in the socio-economy'* (Martin & Sunley, 2006, p. 7). The best-acknowledged distinction in literature results from Liebowitz and Margolis (1995), who found three degrees of path dependency: no historical influences on future economic incidences, a relative influence or a very significant role of path dependency. Interestingly, the third degree (highest level of path dependency) resulted in negative economic growth due to the presence of a full lock-in – the resistance to alter to superior alternatives due to historical developments. However, these degrees are based on static criteria while a global implementation for path dependency will be dependent on dynamic criteria. This makes it improbable to distinguish between strengths of path dependency on a global case, but applicable to specific regions or cases (Martin & Sunley, 2006).

Secondly, there is a constant debate in the scientific literature about the main sources of regional path dependence. Natural resources may be the main reason for early dependence and attraction of homogeneous firms; cultural traditions will foster a local trajectory; and a city's infrastructure may only be feasible for a specific industry, resulting in different sources of path dependency (Martin & Sunley, 2006). Nevertheless, a clear consensus can be seen between studies about the role of traded and untraded interdependencies, influenced by externalities and path dependency. The presence of *'Economies of agglomeration'*, *'Local external economies of industrial specialisation'* and '*Inter-regional linkages and inter-dependencies'* are main sources of path dependency, resulting in several externalities to co-exist and co-operate within a local economy (Martin & Sunley, 2006, p. 16). Thus, regional economic integration generates a stronger sense for path dependency to be present as a result of integrated activities.

Thirdly, debates about path dependency often use several terminologies and notions to what extend path dependency is applicable. This is related to the ambiguous use of the term 'historicity' to indicate path dependency. Historicity, which can often be found in biological, revolutionary and anthropology studies, is repeatedly used as a direct interpretation for path dependency (Ereshefsky, 2014; Martin, 2011; Martin & Sunley, 2006). However, looking at the different interpretations of historicity in other studies, it can be concluded that the terms historicity and path dependency are not identical (Desjardins, 2011). Where both notions focus on how circumstances and incidents in the past affect current and future outcomes, historicity focusses only on the initial conditions for a future outcome to occur. Nevertheless, path dependency theory includes all relevant situations, circumstances and conditions, that can affect the outcome along the way. This is also described by Ereshefsky who states that in the case of *'path dependency, not only do initial conditions affect the probability of an outcome but so do events along the path, as well as the order of those events'* (2014, p. 716). Therefore, path dependency includes a dynamic and multi-dimensional view on the occurrence of externalities within a regional economy.

Based on these notions it can be stated that path dependency is affected by multiple dimensions and economic levels. They generate a local economy that is highly integrated as a result of historical occurrences and dynamic developments. This cyclical process is also captured in figure 2.1. This figure shows the interaction between different economic levels within the society and the interrelatedness of local economic environments. So, path dependency shapes the local economy and affects to what degree regional externalities occur (Press, 2006).

Agglomeration level	*Diversification of economic activities and organisations enriched by Jacobs' externalities*	*Specialization of economic activities and organisations enriched by MAR externalities*
Macro-industry level	*Cultural and economic activities of individuals and organisations contributing to the maintenance or change of cluster-level factors*	
Meso-industry level	*Organisations of individuals, e.g. firms, research institutions and government agencies*	
Micro-industry level	*Individuals and their behaviour*	

Fig. 2.1. *Path dependent economic interaction of macro-, meso- and micro-level industry and agglomeration economies. Reprinted and adapted from Press (2006, p. 63).*

Dependent on the regional scope of the agglomeration it may concern cities within a province, industrial sectors within a metropolis, or major agglomerations within a state. As the following chapter will elaborate, in this study it concerns major Chinese cities within a regionally connected economy. Therefore, from now on this study will mention 'city' when indicating a node or entity within the agglomeration.

2.5 Externality interdependencies

Before the following section elaborates the specific implementation of path dependency theory on the presence of externalities, the different notions of economic interrelatedness, integration and interdependencies need to be amplified. These notions and their implementation in geographic studies are often inconsistent. To avoid misunderstandings, this study enforces the widely acknowledged clarifications of Feld and Boyd (2006). They state that if the ease of performing trade and business between different economies becomes easier, both economies create a sense of economic integration. As a result of this enhanced economic integration, both cities benefit by means of competitiveness, consumption and exports. This will create a certain degree of interdependency on several aspects like transport, infrastructure and trade. It is important to note that interdependency is present once its' disappearance will have a negative effect on the particular economy.

In this study, an increased integration of economies will foster the occurrence of interdependencies between cities within an agglomeration as a result of the supportive presence of externalities. Consecutively, integration is seen as the notion of economic converging towards each other, while interdependencies are present when an economy experiences regional integration (Feld & Boyd, 2006).

Having a clear distinction between economic integration and interdependencies, economic interrelatedness has additional ambiguous explanations. Nevertheless, economic interrelatedness between various cities within a spatial agglomeration has received a growing awareness in the last two decades (Burger & Meijers, 2016). The economic interrelatedness between two cities determines the degree of economic integration, as the result of resemblances in economic structures. This results in a theoretical interconnected framework between economic interrelatedness, integration and interdependencies, shown in *fig. 2.2*.

So, as discussed by Press (2006), a local economic structure is a result of a path dependent process, affected by multiple dimensions and historical developments. When taking into account agglomeration economies, the interrelatedness and economic integration between two cities is a dynamic process. As a result of economic integration, urban interdependencies are created.

However, various studies have tried to capture the different means for the results of economic integration and interrelatedness (Burger & Meijers, 2016). This is also reflected in the wide-ranging theoretical terminology for intra-agglomeration interdependencies, such as *borrowed size* (Burger & Meijers, 2016; Phelps, Fallon & Williams, 2001), *agglomeration externality fields* (Phelps, 1992) and *urban network externalities* (Capello, 2000).

However, whereas some of these studies focus on the economic structures and industries of cities within the agglomeration (Audretsch & Feldman, 2004; Burger & Meijers, 2016; Pessoa, 2011), or adapt an institutionalist perspective on agglomeration development (Amin, 1999; Hassink, 2005), the dynamic process of economic interrelatedness within the agglomeration is neglected.

***Fig. 2.2**. Theoretic framework for externality interdependencies.*

Figure 2.2 captures these relationships between path dependency, economic interrelatedness, integration, and interdependencies, and economic externalities. In summary, intra-agglomeration interrelatedness and integration are caused by path dependent happenings and influences. This results in a certain level of economic interdependencies between cities within the agglomeration. And as earlier elaborated, the presence of economic integration and interdependencies is in line with the collective presence of Marshall-Arrow-Romer and Jacobs' externalities, which fosters intra-agglomeration economic growth. However, whereas economic interdependencies often refer to other (mostly national) trade deficits, and neglects the path dependent and dynamic process of economic integration, this study will implement the notion of '*externality interdependencies'*. Where 'externality' refers to the complementary role of externalities in general, 'interdependencies' refers to the path dependent aspects of economic integration. Thus, the notion of externality interdependencies captures the dynamic role of externalities within agglomeration, whereas the dynamic role is a function of historical developments (path dependency).

Previous literature shows that the dynamics of path dependency, and thus the presence of externality interdependencies between two cities, is effect by several concepts. The following section will analyse these factors and theoretically state their hypothesized effect on the presence of externality interdependencies. Due to the introduction of externality interdependencies, the intertwined relation between these concepts can be theorized and empirically verified. These concepts encompass aspects from different studies related to economic integration and agglomeration studies and include endogenous economic growth, institutional distance, spatial proximity and urban similarities of industrial structure.

2.5.1 Endogenous economic growth

As path dependency states, the dynamics of local integration are contingent on the economic development within the agglomeration. Several endogenous growth models seize the relationship between economic growth and the presence of agglomeration externalities (Klenow & Rodriquez-Clare, 2005; Romer, 1986)

Romer (1986) explains externalities as the stock of knowledge, human and physical capital available in the regional economy. He states in his AK-model (see *eq. 2.1, 2.2* and *2.3)* that each firm within the agglomeration benefits of this stock of knowledge capital, generating increased productivity and innovation capabilities. This results in enhanced economic growth as a result of these externalities, creating a correlated circle for an endogenous growth pattern (Klenow & Rodriquez-Clare, 2005). So, '*Human capital accumulation, in turn, facilitates the introduction of higher quality goods, which are intensive in human capital'* (Klenow & Rodriquez-Clare, 2005, p. 4). In addition to Romer's (1986) human capital externalities, Howitt and Aghion (1998) elaborate externalities in their endogenous growth model as the technology frontier which diminishes due to the occurrence of technological externalities. They state that their growth model implies that the incentive to accumulate capital is the main determinant for regional firms to perform R&D. Due to an increased presence of R&D facilities, the technological frontier within the agglomeration reduces as a result of technological externalities. A theoretical analysis shows that these models are mainly based on the theory of MAR externalities.

In addition, the interrelatedness between cities increases if it experiences endogenous economic growth. This is the case as economic growth and economic size are directly positively related (Chen, 2002). So, encouraged endogenous economic growth '*mostly correspond to the larger-sized urban areas, which would consequently encourage the growth of cities for various reasons, such as scale economies, shared inputs and reduced transaction costs, etc'* (Chen, 2002, p. 532). This results in an ongoing supportive circle between externalities promoting economic growth, and economic growth-enhancing the presence of externalities. Generally, the larger the economy of a city, the more interaction and externalities are present.

Regarding this interaction, De Groot et al. (2009) investigated the importance of time between economic growth and externalities. In their review paper, they examine the theoretical background of 31 related scientific articles on externalities and urban growth. Besides the conclusion that externalities in general favor urban growth, they studied the role of time dimensions. They state that '*it turns out that using cross-section techniques considering longer time periods tends to increase the chances of finding significantly positive effects'* (De Groot et al., 2009, p. 19). Externalities within agglomerations are not present directly and can vary between a subset of channels and dimensions before it can be empirically stated as economic growth. Thus, to appropriately measure the interaction between economic growth and externalities a time dimension needs to be incorporated.

However, it is noteworthy that economic growth is frequently connected to substantial pecuniary diseconomies. These diseconomies, like increased wages and renting prices, traffic crowdedness, pollution and other relevant population-related issues, arise often in the core cities. This is also indicated by Richardson (1995), who states that due to this increased spatial concentration the structure of the agglomeration becomes more extended. As a result '*the economies begin to be overwhelmed by the congestion effects, firms (and households) can escape the consequences by relocating to decentralized locations'* (Richardson, 1995, p. 123). On the one hand, this results in additional flows of externalities within the agglomeration and belittles the pecuniary diseconomies for the core city. But on the other hand, pecuniary diseconomies will diminish the economic integration between cities, effecting the presence of externalities. In line with the general belief within geography studies, it is expected that economic growth positively influences the presence of externalities.

Based on these theories it can be stated that economic growth results in increased externalities over time. This theoretical analysis generates the following hypothesis:

Hypothesis 1: Endogenous economic growth and externality interdependencies have a positive correlation.

2.5.2 Institutional context

Institutional economics has introduced the belief that collective forces, established on micro-, meso-, and macro-industry level, shape the economic environment over time. Where the concept of institutions has received several interpretations and influences, this study clarifies institutions as *'humanly devised constraints that structure political, economic, and social interaction* ... [and that] *provide the incentive structure of an economy'* (North, 1991, p. 97). When taking into account the institutional context within the agglomeration, the influence of path-dependency is even more prevalent. This is also elaborated by Amin (1999):

> *These forces include formal institutions such as rules, laws and organizations, as well as informal or tacit institutions such as individual habits, group routines and social norms and values. All of these institutions provide stability in the real economic context ... and guiding individual action. ...* . [This institutionalist perspective] ... *forces recognition of the path- and context-dependent nature of economic life, or, from a governance perspective, the wide field of institutions beyond markets, firms and states.* (p. 367)

The social emergence of formal and informal institutions is often a slow and evolutionary process. However, once these cultural, social or institutional ties or internal values have been created, it can enhance economic relations between cities or entities with similar institutional frameworks. Cities which experience well-established arrangements and structures that show resemblances to other cities' institutional frameworks experience a greater strength of 'relation assets' (Storper, 1997). Whereas relational assets directly impact the urban network, it affects the degree for externalities to occur. These assets establish trust and reciprocity due to the institutional resemblance; amplify the ease for conducting business with each other as a result of similar (commercial) habits and norms; and a superior common understanding is generated due to the same means of communication and interaction (Amin, 1999; Storper, 1997).

Institutions and their relevant relational assets are present in all layers of socially. Since they are socially constructed the socio-cultural environment is often the main historical determinant. On an international network-level most (governance) institutions are formed based on a neoliberal ideology, as globally this ideology and the free-market economy are dominant (Dicken, 2015). This makes the forms of collaboration and interaction the most desirable one (for instance UN, IMF, World Bank, WTO). On the national level these institutions are often defined by governmental policies, generating a pleasant environment to attract (international) investments and organizations, stimulate social equality or concerning the development of less-favoured regions (Storper, 2004). However, it is important to note that these institutions are not territorially bounded; deterritorialized institutions present in several international markets are a result of the ongoing drive towards globalization (Storper, 2004). Thus, it involves several cities or districts with relational assets based on their geographical perspective within the agglomeration. This path dependent connection between geographical proximity and institutional distance will be discussed more broadly in the following subsection regarding spatial connectivity.

Studies have explored the relationship between the institutional environment in agglomerations and its relation to economic growth or development (Kumssa & Mbeche, 2004; Pike et al., 2015). However, the interaction between institutions and geographical areas is ambiguous. Pike et al. (2015) state that '*the causal linkages and interactions between institutions and economic development at different geographical scales remain unresolved. Whether they exist, their extent and nature, and in what directions they operate are unclear'* (p. 187). Consequently, different hypothesized relationships are present. Kumssa and Mbeche (2004) stated that institutions in several parts of Africa have a weak or even ineffective effect due to the poor enforcement of regional policies, securities and a certain level of corruption and mismanagement, which is in line dependent on the national or regional structure of the institutions. Pike et al. (2015) investigated the local economic development of 39 regions in England from an institutionalist perspective. They concluded that local institutions '*are situated within multi-agent and multi-scalar institutional environments and arrangements*' (Pike et al., 2015, p. 202), which makes it impractical to generalize any specific correlations. Therefore, a focus on the specific correlation between a city's institutional framework and its externalities can lead to spatial and path-dependent results, which might result in biased relations to a city's externalities.

Consequently, in this study, it is more suitable to consider the 'institutional distance' between entities, which defines the extent to which institutional environments differ from each other (Bae & Salomon, 2010). Numerous studies have investigated the role of institutional distance on mutual economic flows between countries or states (Bae & Solomon, 2010; Jones & Lundan, 2001). Both studies contribute to the overall consensus that a greater degree of institutional distance results in diminishing economic flows, like FDI and labour flows. For instance, Jones and Lundan (2011) found that former colonies of the United Kingdom still benefit from greater FDI flows due to similar institutional environments.

In addition, Bae and Solomon (2010) state that institutional distance is complementary to cultural distance, indicating a negative effect on economic interaction. This is also stated by Xu and Shenkar (2002), who state that cultural distance is one of the aspects of institutional distance since it is a result of different institutional systems and environments. Vice versa, cultural distance is no applicable measure for institutional distance as cultural distance '*does not capture the complexity of cross-country differences; in particular, it neglects the critical role of societal institutions in articulating, disseminating, and arbitrating cultural and social cues'* (Xu & Shenkar 2002, p. 608). Therefore, the role of cultural differences on the occurrence of externalities is part of the measure of institutional distance.

These theories regarding institutional distance can be associated with Storper's (1997) concept of relational assets. In contrast to the presence of relational assets, the absence of relational assets is a greater barrier for the occurrence of externalities. If two entities or a particular sector within an agglomeration faces different shared values and cultures, it might result in aversion or even antagonism between cities (Storper, 1997). These dissimilarities are a result of various institutional frameworks, which is then reflected in a greater degree of institutional distance (Bae & Solomon, 2010). Thus, a greater degree of institutional distance diminishes the presence of relational assets, reducing externality effects within the agglomeration.

Based on these theories the following can be hypothesized:

Hypothesis 2: Greater institutional distance between cities within an agglomeration is negatively correlated with the presence of bilateral externality interdependencies.

2.5.3 Geographical proximity versus spatial connectivity

It is commonly acknowledged that geographical proximity between nodes in an agglomeration results in a greater transmission of externalities between entities (Audretsch & Feldman, 2004; Burger & Meijers, 2016; Cappelo, 2000; Gerritse & Arribas-Bel, 2018; Pessoa, 2011). Therefore, several studies have focused on the specific geographical transmission of externalities within the agglomeration.

Phelps (1992) and Cappelo (2000) both recognize the presence of agglomeration benefits due to externalities on a spatial dimension. However, they differ in the theoretical approach of the spatial transfer of externalities between each city/node within the agglomeration. According to Phelps (1992), externalities are on a spatial scale zonal and are exclusively dependently on geographical proximity. He states that there are two constraints for geographical externalities to occur: transportation and information costs. While taking into account the diminishing role for transportation and transaction costs, economic activities have become more information intensive over time. An increase in communication or connectivity does not foster an increase in externalities as it only transmits while being geographically close. This results in geographical proximity being a straight barrier for external economies over time, generating solely zonal externalities.

On the contrary, Cappelo (2000) states that externalities are nodal and dependently on the degree of physical connectivity within the agglomeration. Since the externalities are nodal, this theory assumes externalities to arise and transmit from nodes within the agglomeration, compared to regionally dispersed (zonal) externalities. Furthermore, if the connectivity between members of the network decreases, local productivity improves in a downward sloping curve. In addition and specifically related to Jacobs' externalities, she also concludes that '*although cities enter the network with different goals and behave within the network in different ways, co-operative behaviour guarantees the achievement of the most strategic advantages'* (Capello, 2000, p. 1941). In contribution to the belief of nodal externalities, the theory of *borrowed size* indicates that if a city has greater productivity compared to other cities with the same size, it experiences 'borrowed' productivity of major nearby cities (Phelps et al., 2001).

This is also in line with findings of Gerritse and Arribas-Bel (2018). They empirically find that *'efficient infrastructure, such as a good road network, increases the effective proximity of citizens, and should increase the benefits of population agglomeration'* (Gerritse & Arribas-Bel, 2018, p. 1146). The ease of travelling enhances the urban network, which allows for externalities to occur between previously geographical dispersed cities. Interestingly, Gerritse and Arribas-Bel (2018) also state that besides enhanced externalities that occur explicitly due to an increased connectivity between cities, more efficient infrastructure can also foster a change in transport mode and social interactions.

Nevertheless, it should also be noted that an increase in connectivity within an urban agglomeration can have a negative effect on agglomeration externalities. As urban connectivity increases, externalities might foster competitiveness between homogeneous industries/cities (Burger & Meijers, 2016). This might result in unequal profitability of the increased connectivity due to different levels of urban network embeddedness within the agglomeration. However, as Jacobs' externalities might decrease due to the increased competitiveness, spatial proximity of homogeneous industries fosters economic interrelatedness based on MAR externalities. Besides, it is believed that an unequal profitability enhances the expected positive effect of spatial proximity on externality interdependence to a certain degree.

Whereas the abundance of literature states that connectivity is the main determinant for improved externality occurrence, Phelps (1992) notion of zonal externalities has received minor scientific support. In addition, Phelps' statement that information intensity rules out the diminishing role of transportation costs is outdated, as technology has fostered several new means of communication/transportation between cities and in general since the early 1990s.

Based on this notion the following hypothesis can be formulated:

Hypothesis 3: Spatial connectivity between cities within an agglomeration is positively associated with externality interdependencies.

Besides this unilateral effect of spatial connectivity, an interaction of spatial connectivity and institutional distance is present within the literature. As theorized before, cities that have similar cultural, social or institutional ties and values tend to have greater economic relations between cities (Storper, 1997). Storper's 'relation assets' are based on historical developments, market ideologies or other regional historical determinants. However, path dependency states that these developments often have a regional influence or embeddedness, generating a certain sense of regional institutional similarities. This is also highlighted by Gertler (2005). He studies factors influencing local endogenous growth and states that, from a path dependent perspective, institutions can be described *'as national or regional 'production cultures' or 'embeddedness' can be thought of as a real and significant component of 'industrial history...literally embodied in the present''* (Gertler, 2005, p. 25). This embeddedness is a practice of historical occurrences and thus associated with particular regions. Therefore, it is theorized that spatial connective and institutional distance have a greater complementary role on externality interdependencies, compared to other specified concepts. This results in the following hypothesis:

Hypothesis 4: Spatial connectivity and smaller institutional distance between cities in an agglomeration are highly complementary towards externality interdependencies.

2.5.4 Economic industry structure

As mentioned before MAR and Jacobs' externalities generate supportive externalities, as a result of dynamic economic interrelatedness. Previous studies have tried to capture which externality is superior, but this generated ambiguous results (Di Clemente, Chiarotti, Cristelli, Tacchella & Pietronero, 2014; Ejermo, 2005). Nevertheless, due to the introduction and incorporation of path dependency, theoretical findings on this supportive relationship indicate a historical interrelatedness between cities based on their industry activity. Di Clemente et al. (2014) found results that a higher degree of specialization is the major determinant for microlevel competitiveness, while diversification is the main cause for macrolevel competitiveness. They speculate that '*the diversification success of a firm is the result of the evolution in a competitive environment'* (Di Clemente et al., 2014, p.4).

Contributing to this, Malizia and Feser (1999) state that the formation of an agglomeration starts due to the creation of a specific attractive economic environment. This is caused by the presence of localization economies, and thus specialization. Based on these specialized environments, other related and unrelated industries within the agglomeration profit from the enhanced economic conditions by means of MAR externalities, which eventually generates Jacobs' externalities and a sense of diversification (Malizia & Feser, 1999).

Empirically, Shihe, Xiaofend and Guojon (2010) find that industry diversity is caused by industry churning. A higher degree of industry churning, which is *'the fluctuation or turnover of* *industries in a city'* (Shihe et al., 2010, p.3), is caused by an increased attractiveness of the local economic environment. Firstly, this enhanced attractiveness is caused by the presence of MAR externalities arriving from a dominating homogeneous industry. By attracting industries from other sectors, the level of industry churning and labour turnover will increase. So, secondly, Jacobs' externalities will become more relevant, resulting in a higher degree of regional diversification. According to these theories, industry similarity (localization economies) between two cities at an early stage contributes more to externality interdependency, compared to industry diversification.

Nonetheless, it should be noted that at a given point industry similarity has reached a point whereas localization effects are not significant anymore (Hong & Xiao, 2016). This is due to the significant rise of Jacobs' externalities. Given the presence of a free flow of externalities and openness, this given point should be relatively invariable within the agglomeration. This is as MAR externalities are beneficial to the absolute agglomeration by means of potential and attractiveness, and not limited within a city (Guiliani, 2003).

Based on these theoretical and empirical findings it can be hypothesized that an early level of specialization, and thus industry similarity, has a significant effect on the total presence of externality interdependencies. Whereas the literature does not indicate any significant complementary relations between industry similarity and either spatial connectivity or institutional distance, no hypothesis is suggested for these intertwined concepts. This result in the formation of the following hypothesis:

Hypothesis 5: Industry similarity between cities is positively related to the presence of externality interdependencies.

2.6 Econometric model

Based on the theoretic framework and the path dependent approach on externalities, the phenomenon of externality interdependencies needs to be econometrically formulated. Since this situation involves a set of relationships that explain the dynamic and simultaneous presence of externalities within the agglomeration, a simultaneous equations model needs to formulated (Shalabh, 2012). This model allows to formulate two equations which capture all externality interdependencies since both equations are integrated.

Before the econometric model can be formulated, the variables in simultaneous equation models need to be classified as endogenous (also described as jointly determined) or exogenous (Shalabh, 2012). Where the former relates to a variable that is determined by values or functions of relations in the model, the latter is predetermined outside the model. Noteworthy is the fact that in simultaneous equation models, variables can have different classifications in different equations. So, in line with the assumptions of endogenous economic growth in agglomeration economies, economic growth is the main variable that is determined in the model and is an endogenous variable. The other relevant factors in the agglomeration effect endogenous economic growth and are thus exogeneous variables. It is noteworthy that endogenous economic growth and endogenous variables are not similar presumptions since the first represents an economic theory which states that economic growth is caused by internal factors, and the prior describes the statistical role of a variable within the simultaneous econometric model.

Now, the endogenous simultaneous econometric model can be formulated and written as

$$y_{i,t+1} = \sum_{j=1}^{n} X_{ji,t} + \beta_2 C_{i,t} + u_{i,t} \qquad (2.4),$$

in which

$$X_{ji,t} = y_{j,t}\left(S_{ji}, ID_{ji}, I_{ji}, C_j\right) \qquad (2.5).$$

Let *i* and *j* denote the city of relevance, where *j* is limited to *n* cities in the agglomeration, at time *t*. In (*2.4*), economic growth (y) is dependent on the sum of all externalities (*X*) from city *j* to city *i*, as well as controls (*C*) and an error (*u*). The second equation (*2.5*) captures externalities from city *j* to city *i* as an endogenous factor, which is dependent on the spatial connectivity (*S*), institutional distance (*ID*), and industry similarity (*I*) between city *i* and *j*, and controls (*C*) for *j*, as a factor of economic growth.

As a function of the simultaneous models (*2.4*) and (*2.5*), the specific externalities flowing from one city to another city within the agglomeration can be analysed. As stated by Qureshi and Compeau (2009), '*the ability to detect the presence or absence of between-group differences and accurately estimate the strength of moderating effects is important in studies that attempt to show contingent effects'* (p. 197). In addition, the path dependent approach allows analysing the effect of certain circumstances or developments on the economic interaction due to time variations. Thus, outlining receiving or outgoing externality interdependencies allows for several implications, for instance, core-periphery studies, indirect economic effects and regional policy studies. Obviously, within an agglomeration, the effect of externalities between certain cities can be insignificant. Lack of connectivity, great institutional distance or any other determinants can cause this non-integration. In order to analyse these underlying causations, eq. (*2.6a*) and (*2.*6b) are formulated:

$$\sum_{j=1}^{n} X_{ji} = \left(X_{1i,t}, X_{2i,t}, X_{3i,t} \dots X_{ni,t}\right), \qquad j = 1,2,3, \dots n \qquad (2.6a).$$

$$\sum_{j=1}^{n} X_{ij} = \left(X_{i1,t}, X_{i2,t}, X_{i3,t} \dots X_{in,t}\right), \qquad j = 1,2,3, \dots n \qquad (2.6b).$$

Depending on the city of interest, *i* represents in (*2.6a*) the receiving city, while *j* represents all other cities present within the agglomeration. Vice versa, to measure the sum of outward-bound externality interdependencies eq. (*2.6b*) needs to be adapted, which focusses on the flow of externality interdependencies from city *i* to the other cities (*j*) within the agglomeration.

3. Case study: Greater Pearl River Delta

To empirically test externality interdependencies an agglomeration needs to be included for analysis. As highlighted previously, a path dependent approach reduces the possibility of a highly desired objective in scientific studies, namely generalization of results. This relates to the unique path of historical interactions and developments, generated by (often) singular circumstances. Therefore, while conducting an agglomeration study from a path dependent approach, the agglomeration of study needs to be carefully analysed and extensively elaborated in order to theoretically demonstrate any internal paths or developments. The next section elaborates the development of the Greater Pearl River Delta, as well as the (economic) position of each city within the agglomeration. Secondly, a comparison of this agglomeration to related agglomerations specifies the uniqueness and variations amongst them.

The Pearl River Delta (PRD) relates to a jurisdiction of Guangdong Province in the southern part of the People's Republic of China (PRC), compromising nine major prefecture-level cities: Dongguan (Do), Foshan (Fo), Guangzhou (Gu), Huizhou (Hu), Jiangmen (Ji), Shenzhen (Sh), Zhaoqing (Zha), Zhongshan (Zho) and Zhuhai (Zhu). It derives its name from the Pearl River, the third-longest river of the PRC, which passes Guangdong Province and eventually flows into the South China Sea. Located on a strategic position, the PRD has favoured geographical advantages of functioning as the gateway to the hinterland of mainland China and within spatial proximity to neighbouring economic strongholds, like Singapore, Shanghai and Kuala Lumpur (Bie, de Jong, & Derudder, 2015). As a result, this area has been one of the greatest contributors to the tremendous economic growth of the PRC since 1980 and marked South-East China as a weighty economic centre worldwide. Geographically close to Guangdong Province are the jurisdictions of Hong Kong Special Administrative Region (HKSAR) and Macau Special Administrative Region (MSAR) (see figure 3.1). If HKSAR and MSAR are included in the definition, the area is expressed as the Greater Pearl River Delta (GPRD).

Currently, the GPRD has a combined population of over 70 million, 12% of the total GDP of the PRC, three of the world's top ten containers ports, two stock exchanges, and is home to several political, cultural, financial, transport and economic centres (Tang & Ellison, 2019). The tremendous economic developments within the GPRD have established a great agglomeration of interdependencies and economic integration, contributing to the vast economic growth within the region.

Fig. 3.1. *Map of Greater Pearl River Delta, HKSAR and MSAR. Reprinted from Tang & Ellison (2019).*

These developments within the Greater Pearl River Delta, compromising two former European colonies and several economic zones, have resulted in a global multi-industry economic hub. Based on the theory of path dependency it is important to elaborate these intra-agglomeration relations and economic developments. Thus, this chapter does not seek to elaborate each economic development and its significance to the evolution of intra-agglomeration externalities. Instead, different historical and economic characteristics are emphasized in order to interpret disparities between inter-city externalities. As this analysis focusses on the economic structure of the GPRD before the implementation of the Development Plan for the Greater Bay Area, it includes significant policies and developments before 2019.

The following section briefly elaborates the historical development of the GPRD, the nine prefecture-level cities and the two Special Administrative Regions. Furthermore, the formation of other economically advanced urban agglomerations in the world is given in order to give insights into the non-uniformity of the formation of urban agglomerations.

3.1 Regional and historical developments

Due to rapid economic developments during the past three decades, the GPRD '*has become one of China's three main economic hubs and reputedly a 'world factory''* (Bie et al., 2015, p. 104) – next to the Yangtze River Delta including Shanghai and the Jingjinji Metropolitan Region surrounding Beijing. The GRPD, which was preliminary a rural community 40 years ago, has now become one of the most urbanized, densely and largest urban agglomerations of the world. Nowadays, the region has a population larger than the UK or France, and a total GDP larger than that of Australia. These (economic) characteristics of the GPRD and its cities are shown in table 3.1. Obviously, globalization and the increased openness of mainland China have contributed to these developments significantly. This has resulted in a highly integrated regional economy, complementing each other's industries and promoting overall city accommodations, like infrastructural improvements and intra-agglomeration science parks (Bie et al., 2015).

Nevertheless, the GPRD is also unquestionably divided, consisting of different regimes, economies, and industrial structures due to historical segregation's and interventions (Vogel et al., 2010). Even nowadays, the GPRD faces three different monetary, tax and legal systems; two Special Economic Zones; three competing Free Trade Zones; and several highly competing transportation facilities. These variations are also mentioned by Khanna and Sevcik (2012), who described the following:

> *In a single day, you can cross from Hong Kong, an open society with aggressively free media, to more state-directed but still very global Shenzhen or less-glitzy Dongguan, to the freewheeling and somewhat sleazy gambling haven of Macau, or the tax-free master-planned Zhuhai/Hengqin. Along the way, you will go through checkpoints ranging from full-fledged border crossings requiring visas to light security checks. The journey reveals the different constitutions and political priorities playing out even as the Delta region physically becomes ever more one single city.* (para. 4)

The following section elaborates any historical influences on the position of each city within the Greater Pearl River Delta, its current economic structure and other relevant events related to a path dependent approach.

	GDP (2018, CNY 100 million)[1,2]	*GDP (2000, CNY 100 million)[1,2]*	*GDP per capita (2018, CNY)[1,2]*	*Population (2018, mn)*	*Land Area (sq. km)*	*Primary Industry (2018, % of GDP)*	*Secundairy Industry (2018, % of GDP)*	*TeritaryIndustry (2018, % of GDP)*
GPRD	**110,133**	**20,841**	**130,182**	**71.16**	**56,904**	**-**	**-**	**-**
HKSAR	25,249	11,952	337,553	7.48	1,107	0.1	7.5	92.4
MSAR	3,835	468	572,388	0.67	33	0	5.1	94.9
Gu	22,859	2,493	155,491	14.9	7,434	1	27.3	71.7
Sh	24,222	2,187	189,568	13.03	1,997	0.1	41.4	58.8
Fo	9,936	1,050	127,691	7.91	3,798	1.5	56.5	42
Do	8,279	820	98,939	8.39	2,460	0.3	48.6	51,5
Hu	4,103	439	85,418	4.83	11,347	4.3	52.7	43
Zho	3,633	345	110,585	3.31	1,784	1.7	49	49.3
Ji	2,900	505	63,328	4.6	9,507	7	48.5	44.5
Zhu	2,915	331	159,428	1.89	1,736	1.7	49.2	49.1
Zha	2,201	250	53,267	4.15	14,891	15.8	35.2	49

Note. [1] *At current market prices.*
[2] *Computed using yearly average exchange rates.*

Table 3.1. *Major economic indicators of the Greater Pearl River Delta (2018). Data from HKTDC Research (2019) and from Guangdong Statistical Yearbook (Guangdong Statistical Bureau (GSB), various years).*

Hong Kong SAR (HKSAR)

Hong Kong has been a colony of Great Britain since the Treaty of Nanking in 1842 until the handover of Hong Kong to the PRC in 1997. During these 156 years of colonisation the British have controlled Hong Kong as an open market, applying British rights and laws, spreading the English language and allowing a relatively high degree of autonomy (Vogel et al., 2010). As part of the handover to the PRC, the metropolis became the Hong Kong Special Administrative Region (HKSAR), including the 'one country, two systems' (OCTS) agreement in return for the handover. The OCTS includes the right for the HKSAR to enjoy a high degree of autonomy, and mostly maintain their right to keep their colonial legal system, its democratic governance and its economic structure (Bie et al., 2015). Therefore, HKSAR, while being part of China, has its own legislative council which controls and executes the Basic Law (constitutional framework specifically for HKSAR). This Basic Law states that HKSAR stays in control of their own affairs, except for PRC's foreign and defence affairs. Besides its legal system, a different monetary and tax system is present. Instead of the Chinese Yuen, HKSAR utilizes the Hong Kong Dollar, and tax rates are significantly lower compared to mainland China (Bie et al., 2015; Vogel et al. 2010).

Nevertheless, the handover of HK to the PRC also resulted in increased regional economic interrelatedness. Due to increased accessibility between people from mainland China and HKSAR, interregional connections were established, including a highly anticipated high-speed train network in 2018. This was the result of renewed efforts to stimulate regional economic ties by implementing '*a series of intergovernmental agreements and mechanisms between Hong Kong and Guangdong province in 2003. …* [Both cities] *set up a framework for the two governments to meet regularly and establish common policies for economic development, infrastructure, tourism and public health'* (Vogel et al., 2010, p. 66). In addition, since HKSAR is considered a 'free trade zone', it has experienced significantly lower import and export rates, attributing to HKSAR's position as a transport hub. This contributed to the position of HKSAR as the financial centre of Asia, since lower tax rates enhance '*Hong Kong's leading role within Asia as an intermediator of FDI flows…'* (Leung, 2008, p. 6), stimulating economic flows between mainland China and HKSAR. Due to this appealing financial environment, HKSAR benefits from a high service economy and almost no manufacturing industries.

Macau SAR (MSAR)

In line with Hong Kong, Macau owns its social, economic and political characteristics to its colonial past. The Portuguese Empire has been using Macau as a trading settlement since 1557 and received colonial rights in 1887. Macau, under Portuguese administration, did not expand immensely (as was the case regarding HKSAR) and its primary function remained trade until its handover to the PRC in 1999 (Hao, 2011). Where Hong Kong enjoyed intense economic growth, Macau was relegated to a quiet port town, operating as Christianity's gateway to China.

The handover also included an agreement that Macau becomes a Special Administrative Region, and operates under the principle of OCTS. Therefore, Macau has a higher degree of autonomy and a European governance structure; has a different legal and monetary system (Macanese pataca) and has an inherited legal gambling industry. In addition, Macau is marked as the only legal gambling place within China and experienced intense economic growth due to the opening of the border. Gambling facilities, infrastructural projects and related industries were promoted to attract regionalism and tourism. This resulted in '*brilliant economic achievements as well as a structural shift toward a gaming-industry dominated* economy' (Sheng & Gu, 2018, p. 72), which resulted in a rapid pace of GDP growth and increased (wealthy) tourism.

Guangzhou (Gu)

The capital and most populous metropolis of the Guangdong province, Guangzhou, experienced tremendous economic growth since the 1980s as a result of the PRC's economic policies. As the previous main port of the Maritime Silk Road, the city historically became the transportation hub in the PRD and the southern gateway for mainland China. In addition, the Chinese Communist Party (CCP), led by Mao Zedong, was founded in Guangzhou. This resulted in Guangzhou becoming a major cultural and political centre while focusing on national and regional connectivity. Therefore, for the PRD, Guangzhou became the main place within the (provincial) transportation network (Hui et al., 2018).

Furthermore, the PRC promoted economic activity in Guangzhou since the Open Door Policy 1980s due to its historical trade environment. The city rapidly attracted regional manufacturing firms, generating a high level of industrialisation. Therefore, Guangzhou is often seen as the historical manufacturing hub in the Pearl River Delta (Hui et al., 2018). This, and the close proximity to HKSAR, have been reasons for the PRC to appoint an area of Guangzhou as a Free Trade Zone in 2015. In these zones, (foreign) companies pay less taxes, face less financial and legal controls, and Chinese customs and import duties are not applicable (Sohlberg, 2018).

Shenzhen (Sh)

By the end of the 1970s, Shenzhen was primarily an agricultural and border town of Hong Kong. However, Shenzhen (together with Zhuhai, Shantou and Xiamen) became one of China's first Special Economic Zones in 1980, resulting in the Shenzhen Special Economic Zone (SSEZ). The SSEZ made local authorities '*entrusted with the task of running a local economy instead of just following investment decisions from central ministries'* (Ng, 2003, p. 431). This resulted in tremendous economic growth (improving 1,152-fold) and population growth (an increase of 4200%) between 1979 and 2017 (Ng, 2003; UN-Habitat, 2019).

By appointing Shenzhen as a SEZ the city became a modernized zone, attracting technological investments and fostering economic linkages. Due to the SSEZ, firms located in Shenzhen were one of the first firms that were obliged to perform international trade activities, received several tax benefits and profited from several (special) policies with regards to economic regulations, organizational authority and international investments (Ng, 2003) . This generated a localized and specialized high-technology cluster, making Shenzhen the R&D, innovation and technology capital of the GPRD and PRC. Nowadays, the SSEZ has the highest level of GDP within the GRPD; annually more patent applications then Germany and South Korea; and the third-busiest container port in the world (UN-Habitat, 2019). In addition to the SEZ, the PRC appointed an area of Shenzhen as a Free Trade Zone in 2015, stimulating trade and foreign (mostly focussed on HKSAR) investments.

Foshan (Fo)

Foshan, the fourth most populous city within the GPRD, was considered a great commercial centre in the 18th century. During the 20th century, the rising competition within the region precluded these economic developments, until the Open Door Policies of the 1980s. Subsequently, Foshan became '*one of the most important manufacturing bases for not only China, but also the world,* [and] *Foshan is characterized by its rural industrialization and peri-urban sprawl'* (Han & Jia, 2017, p. 108).

Nowadays, according to Wang et al. (2017), Foshan is in the top-5 of the largest manufacturing cities of China. The city's main manufacturing areas are automobile assembly, (electronic) machinery, and electric appliances, with a high differentiation between technological developments, and SMEs and multinationals. This results in manufacturing accounting for almost 60% of the city's GDP – the highest percentage amongst all cities (Wang et al., 2017).

Dongguan (Do)

Strategically located between technological and economic strongholds, Shenzhen and Guangzhou, Dongguan also achieved great economic developments at the end of the 20th century. Previously, the city was mainly an agricultural and mountainous area, focussing on (local) food production. However, succeeding the regional economic developments, Dongguan became the region's most attractive manufacturing base for neighbouring cities within the PRD, and foreign countries like Hong Kong and Taiwan. These inflows of FDI are considered to be the main forces behind the export and import oriented manufacturing base of Dongguan (Yang, 2007).

However, due to increased (foreign) competition and attractiveness, Dongguan's manufacturing hub became less attractive. To cope with this decrease and the developments of the SSEZ, local authorities aimed at attracting high-technology investments creating high-tech industry parks and fostering spatial connectivity. As a result, Dongguan '*is China's most competitive and influential industrial robot industry base, as well as a model intelligent manufacturing city'* (Li et al., 2020, p. 10).

Huizhou (Hu)

Huizhou is the second biggest city within the GPRD and bordering Shenzhen, Guangzhou and Dongguan. The city is often called the Gateway of Eastern Guangdong, as all major connections pass through the city. The city was mainly focussed on agriculture until the CCP introduced the 'Huizhou Dayawan Economic and Technological Development Zone' (Huizhou Dayawan ETDZ) in 1993 (HKTDC Research, 2015a). This development zone has expanded in 2006, covering almost 2.5 times its original size and started developing a selected export zone. The ETDZ's focus industries include electronics, energy, logistics and petroleum, which has resulted in the presence of several energy-related multinationals including Shell and China's national oil company CNOOC (HKTDC Research, 2015a).

Zhongshan (Zho)

Due to its proximity to Guangzhou, the city has developed into mainly a manufacturing city. Several initiatives, like the 'Zhongshan National Torch High-Tech Industrial Development Zone', have been established in the 90s, in order to stimulate economic developments. However, neighbouring competition from development zones in Shenzhen and Guangzhou have prevented further specializations within Zhongshan (HKTDC Research, 2015b). In addition, Zhongshan has a limited area to invest in future developments zones which has resulted in the city becoming a major manufacturer for less-technological industries like furniture, lighting, and textile.

Jiangmen (Ji)

Similar to several (relatively) minor cities within the GPRD, Jiangmen (the third biggest city in size of the GPRD) has a vastly developed manufacturing sector due to its proximity to the core cities. In particular, Jiangmen has experienced a high degree of growth in advanced industries. This has been the result of new policies, implemented in line with the 12th Five-Year Plan in 2011. This policy included a new strategy which stated that city '*has three strategic bases for emerging and high-tech industry bases: Electronics and information; New energies; New Materials'* (Deloitte, 2013, p. 10). Nowadays, Jiangmen has 13 specialized industrial parks, of which at least 11 focus on advanced industries, making it a regional hub for advanced manufacturing.

Zhuhai (Zhu)

Together with Shenzhen, Zhuhai was appointed as a Special Economic Zone in 1980. As was the case in Shenzhen, the SEZ's local beneficiaries and policies attracted numerous investments from neighbouring (foreign) Hong Kong and Macau. With several expansions of the SEZ within the city in the 90s, *'the Zhuhai SEZ was further extended to the whole city of Zhuhai … on the National Day of 2010'* (Sheng & Tang, 2013, p. 72). In addition, tax payments have decreased from 25% to 15% in 2010 which further stimulated economic developments within Zhuhai.

Since the city is surrounding Macau, the PRC has appointed a part of Zhuhai as a Free Trade Zone in 2015, similar to Shenzhen and Guangzhou. This FTZ primarily contributes to Zhuhai's connectivity to MSAR and focusses on its main industries electricity, energy, tourism and leisure (Sohlberg, 2018).

Zhaoqing (Zha)

Historically, Zhaoqing was an important city during the Qing dynasty. However, nowadays, the city lost its significant relevance within the region due to the immense competition within the Greater Pearl River Delta. Its main industries are agriculture and basis manufacturing, which has resulted in the city having the lowest GDP per capita in the GPRD.

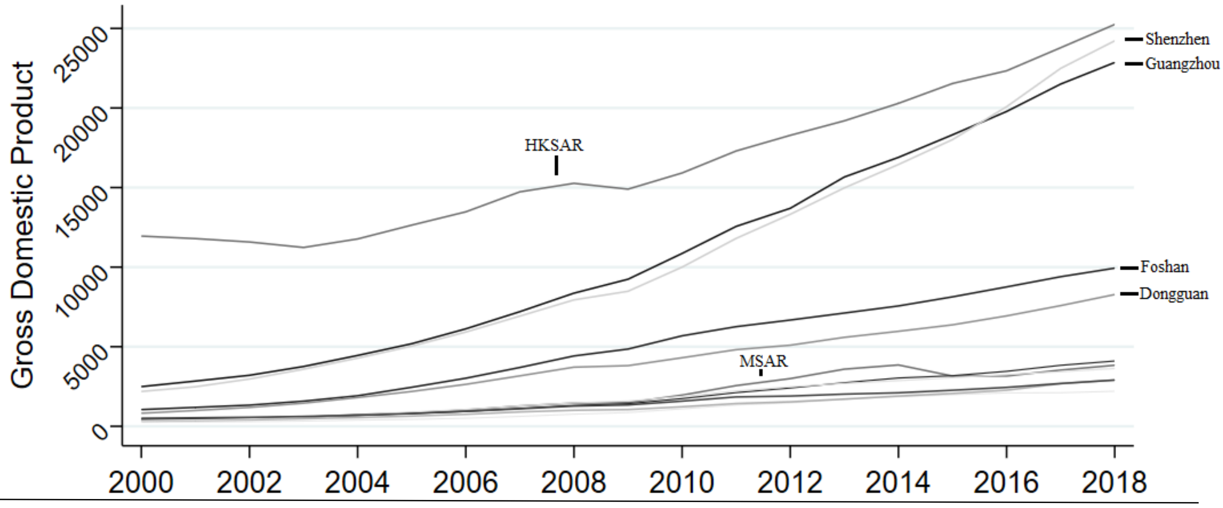

Fig. 3.2. *Gross Domestic Product of the Greater Pearl River Delta from 2000 till 2018. Data retrieved from Guangdong Statistical Yearbook (GSB) (various years).*

3.2 Mega-city agglomerations in the world

As this study produces a nexus of externality interdependencies which can be applied to several agglomerations, it is important to understand similarities and differences between the Greater Pearl River Delta in the PRC and other mega-city agglomerations in the world. This comparison contributes to the previous notion of path-dependency that there is no static economic development direction and all relevant influences are very manifold (Press, 2006). Besides, it shows that the GPRD is exceptional in its historical and institutional developments; varieties in spatial connectivity; and diversified economic structure.

Obviously, a short review of other mega-city agglomerations comes short in pointing out specific differences between each agglomeration. However, any dissimilarities illustrate the uniqueness of each agglomeration and especially the Greater Pearl River Delta. Globally, on an economic level, four mega-city agglomerations are comparable to the GPRD: Jingjinji Metropolitan Region, Yangtze River Delta, San Francisco Bay Area and Greater Tokyo Area. These agglomerations are in economic and social terms relatively comparable to the Greater Pearl River Delta (which are represented in *table 3.2*), as well as a proper representation of various agglomeration formation processes.

Jingjinji (Beijing-Tianjin-Hebei) Metropolitan Region

In terms of area size, the Jingjinji Metropolitan Region is the greatest amongst the five mega-city agglomerations – more than five times the size of the Netherlands. The Jingjinji Region, consisting of the provinces Beijing, Tianjin and Hebei, was first mentioned as a mega-city agglomeration in the 12th Five Year Plan of the CCP in 2011 (Preen, 2018a). Due to its recent and appointed position as an agglomeration, several integration plans are proposed by the government. This includes a travel time of less than an hour between any major city within the Jingjinji Region and 9,000 km of expressways in 2020 (Preen, 2018a).

This newly inaugurated agglomeration surrounds Beijing, the political, educational and cultural heart of the PRC. Tianjin is known for its high value adding industries, and strong manufacturing position in the country. However, the region Hebei, consisting of 11 cities which surround Beijing, is seen as a relatively poor and underdeveloped region. Therefore, the integration plan of the CCP includes the development of multiple R&D centres around Beijing (including the Tianjin FTZ) and promote the local development of value added industries.

Yangtze River Delta

On the other hand, the Yangtze River Delta is naturally created and is now the greatest mega-city agglomeration in terms of GDP (20% of the PRC) and population. It consists of 27 cities of which Shanghai, Nanjing and Hangzhou are the economic strongholds. The region is marked by the Yangtze River, the longest river of Asia. Since Nanjing was the old capital of China, the whole region experiences a high degree of cultural and historical resemblances.

Shanghai has been the economic centre of the PRC since the 1950s, and several surrounding cities were included in 1982 with the establishment of the Shanghai Economic Area. This resulted in vast economic activity within the Delta and promoted surrounding cities as complementary to Shanghai (Preen, 2018b). Shanghai itself is the logistics centre of Eastern Asia, and (due to Chinese capital control) the second financial centre next to HKSAR. Suzhou is the technological manufacturing hub, Nanjing the scientific and education hub, and Hangzhou attracts the highest number of Chinese headquarters. Other nearby cities are specialized in all sorts of manufacturing, natural resources, and attract a large amount of foreign direct investments (Preen, 2018a). The inter-agglomeration cities are connected by one of the most modern transport infrastructures in the world, consisting of several high speed railways, nine international airports and the first and fourth busiest ports in the world (Preen, 2018b). In order to further stimulate the economic development in this Delta, the CCP developed the Shanghai FTZ in 2013.

San Francisco Bay Area

Located in northern California, the San Francisco Bay Area is the smallest mega-city agglomeration in terms of population, size and gross domestic product. However, it has unequivocally the highest GDP per capita (more than twice the GDP per capita of Greater Tokyo). This is the case as the San Francisco Bay Area consists of nine counties, which accommodates after New York the highest concentration of headquarters of Fortune 500 companies (Liu, 2019).

These nine counties include San Francisco, San Jose (Silicon Valley) and the East Bay, which results in the second smallest mega-city agglomeration. Where any historical, cultural or institutional distance between the counties is barely present, each counties' economic activity is clearly divided. Silicon Valley in San Jose has developed into the economic and technological core of the agglomeration, stimulating the role of San Francisco as the financial and tourism hub, and as a locational extension of San Jose (McNeill, 2016). Next to these economic strongholds, East Bay is the manufacturing and distribution county of the agglomeration, like shipping, oil and have industry.

Greater Tokyo Area

The Greater Tokyo Area, also described as the Tokyo Metropolis Region, is made of Tokyo and seven surrounding counties. The whole area has a population of over 36 million, which is almost 30% of the whole population of Japan, and a GDP slightly greater than the GPRD (31.4% of the country). Nowadays, the Greater Tokyo Area is the most populous urban area in the world.

Tokyo's development and position as the capital of Japan only began in the late 19th century, generating a coherent and traditional local culture (Liu, 2019). After the vast destruction of the city during the Second World War, rapid economic developments and population growth followed. Massive investments in transportation facilities resulted in a great level of connectivity within the relatively small Metropolitan Region. Nowadays, the city of Tokyo functions as the financial and managerial core of the urban agglomeration. The seven counties are specialised in strong manufacturing and have a strong export function due to the presence of six major ports in the area.

Five mega-city agglomerations	Population (million)	Area (sq. km)	GDP (billion USD)	GDP per capita (USD)	% GDP of the country	Units
Greater Pearl River Delta	69.6	55,800	1,514	21,764	13%	9 Cities and 2 SARS
Jingjinji Metropolitan Region	112.4	214,900	1,217	10,828	9.8%	13 Cities
Yangtze River Delta	120.7	131,500	2,105	17,428	20%	27 Cities
San Francisco Bay Area	8.8	26,300	863	98,637	4.6%	9 Counties
Greater Tokyo Area	36.3	13,600	1,562	42,865	31.4%	1 City and 7 Counties

Table 3.2. *Greater Pearl River Delta compared to other global mega-city agglomerations (2017)). Adapted from Hongbin and Lam (2018) and Liu (2019, 18 February).*

This short review of similar mega-city agglomerations indicates the exceptionality of path-dependency. Within the PRC, the GPRD is the only agglomeration with former colonial cities, different monetary and legal systems, and two FTZs. On the other hand, the Yangtze River Delta is the most prosperous metropolis, experiencing the greatest level of connectivity and a (theoretical) high level of economic complementarity. On the other hand, the Jingjinji Region is a possible future strategic, political and economic stronghold, but is often criticised for its non-connectedness. The San Francisco and Greater Tokyo Area show two agglomerations which are generated and stimulated due to massive and tactical investments, which resulted in a local network of diversified economies. Therefore, and in addition to the theory, this review indicates that path dependency has a non-static occurrence in this study and is not subjected to generalization.

4. Methodology and operationalization

This chapter provides the operationalization and methodology part of this study. Specifically, the relevant data is collected, the estimation techniques for the multigroup structural equation model are elaborated, and the data is analysed. Firstly, the main research variables will be described, followed by the specific variables for this study. These specific measures include spatial connectivity, institutional distance and industry similarity and are formulated and operationalized appropriately. The following section, estimation techniques, will econometrically specify the formulas, describe the applied methodology and generate the path diagram. Hereafter, the descriptive statistics are used for analysing the data and adapted if necessary.

Panel analysis will be performed in this study, analysing annual data of the eleven cities of the GPRD from 2000 till 2018. Due to the previously elaborated colonial pasts of Macau SAR and Hong Kong SAR and the absence of coherent and consistent databases for the nine cities of the Guangdong province, the year 2000 is an appropriate year to start the analysis in this study. A panel study is choosing as it is the proper method to study the dynamic role of externalities while including several observations simultaneously and incorporating relations amongst the specified concepts. The specific reasoning regarding the multigroup structural equation model follows in 4.2. For data collection purposes as well as the data analysis, STATA software is used. Hereafter, AMOS software is used to formulate and estimated the SEM.

4.1 Data collection

Quantitative panel data will be collected from various databases. Since qualitative data from the specific area of study is hard to collect and can easily be biased due to own interests, quantitative data is the appropriate approach for this study. Nevertheless, quantitative data also raises certain concerns which will be carefully analysed in the discussion.

Data related to the nine prefecture-level cities of the Guangdong province (Guangzhou, Shenzhen, Foshan, Dongguan, Huizhou, Zhongshan, Jiangmen, Zhuhai and Zhaoqing) are retrieved from the Guangdong Statistical Yearbooks (various years), provided by the Guangdong Statistical Bureau (GSB). Data associated with SARs Hong Kong and Macau are collected respectively from various statistical yearbooks by the Census and Statistics Department (2019) of the Government of Hong Kong and the Statistics and Census Department (2019) of the Government of Macau. Other relevant data sources, when applied, are mentioned accordingly. Since monetary data from HKSAR, MSAR and the Guangdong province are often presented in their local currency (respectively Hong Kong dollar, Macanese pataca and Chinese Yuan), data is scaled similarly using yearly average exchange rates.

As mentioned before, previous studies have not analysed the externalities between nodes of an agglomeration using path dependency theories. Since there are no specific equivalent empirical studies, the data collection and variable selection needs to be elaborated extensively. A broad analysis shows that relevant measures for institutional context, spatial connectivity and industry similarity are non-consisting and need to be computed or derived from other datasets.

It is noteworthy that since this study is implementing a multigroup structural equation model in order to compare externality interdependencies between cities of the GPRD, relevant data of these three specified variables are computed on a bilateral scale. This results in values between specific cities for each year, allowing to measure the exact effect of these concepts on the subject of study. These three variables are elaborated broadly hereafter. Firstly, the other major variables are described below. These variables are collected routinely as they do not influence latent structures within the SEM as they are exogeneous factors.

Gross Domestic Product (GDP)

Since this study focusses on the presence of externalities between nodes of an agglomeration, local GDP captures the endogenous economic growth (Chen, 2002; Klenow & Rodriquez-Clare, 2005). While taking into account the proper control variables, GDP growth is the direct result of increased externality occurrence within an agglomeration. Vice versa, Chen (2002) states that if an economy increases its GDP, externalities will have a greater impact on its local economy. Therefore, GDP is used as a measure to capture bilateral externalities, while taking into account proper control variables to measure external influences on GDP growth. In addition, Chen's (2002) theory is in line with the first hypothesis, emphasizing that endogenous economic growth (GDP) is positively related to the effect of bilateral externalities. Thus, GDP at current prices is used as a dependent variable to measure externalities. As elaborated in the theoretical section, the interaction between economic growth and externality interdependencies is subjected to a certain time dimension (De Groot et al., 2009). Therefore, future values of GDP will be incorporated, generating a lead for dependent variable GDP.

Population

Population growth will be included as an independent variable. As described in the theoretical framework, urbanization economies and Jacobs' externalities can be seen as complements and are related (Illy et al., 2009). Therefore, it is expected that population growth is positively related to the occurrence of externalities and thus to GDP.

Wages

Manufacturing wages vary significantly, even regionally. This is, besides MAR and Jacobs' externalities, one of the main motivations for companies and manufacturing firms to relocate or conduct business in a specific region. Regional wage differences can be directly related to externalities since wage differences can be caused by a significant local supply of inputs; or a well-connected region has better networks and market access, resulting in a higher level of attractiveness and greater externalities (Amiti & Cameron, 2007). Therefore, wages are included as an independent variable and is expected to be positively related to externalities. Wages are measured as an average annual nominal wage of all employees per city. For HKSAR, monthly nominal wages are transformed to annual statistics.

Human Capital

Since both MAR- and Jacobs' externalities include the presence of knowledge spillovers, human capital plays a vital role. Human capital is directly positively related and associated with innovation, education and entrepreneurship, which all influence the degree of externalities (Dauth, 2010). Unfortunately, regional education and entrepreneurship measures are absent in data regarding the GPRD. Therefore, human capital is measured as product innovation, using 'output value of new product'. The criteria set by the GSB for 'new products', and translated by Zhang (2015), states that '*new products refer to brand new products produced with new technology and new design, or products that represent noticeable improvement in terms of structure, material, or production process for significantly improving the character or function of the older versions'* (p. 107). Consequently and due to an absence of alternatives, it is an appropriate measure for human capital.

It should be noted that different economic structures (e.g. financial, manufacturing or technological) might bias these measures (Zhang, 2015). While scrutinizing the data, the variable 'output value of new product' indicates an immense human capital growth of Shenzhen during the last ten years; a relatively vast and stable growth of Guangzhou; and a stable, but relatively less developing new product output for HKSAR. These developments are in line with Huifeng's (2018) conclusions. She states that:

> *Shenzhen and Guangzhou have fostered hi-tech industries, like biotechnology, new energy and new materials to spur their economies. Since 2013, Shenzhen has invested more than 4 per cent of its GDP annually on research and development, while Guangzhou says it aims to plough 2.7 per cent of its GDP into R&D this year. In October, Lam* [Carrie Lam, Chief Executive HKSAR] *said Hong Kong would double its expenditure on research and development to 1.5 per cent of its GDP in the next five years to encourage innovation.* (para. 8).

Therefore, it can be concluded that different economic structures do not affect this measure in this case and represent the human capital developments in the GPRD accurately. Similarly, the data indicates a significantly low degree of human capital in Macau SAR. Since Macau's GDP consists mainly of gaming revenues (50%), insurance (10%), and several commercial facilities, innovation is barely represented in the economic development of the SAR.

Foreign Direct Investments (FDI)

As stated by Kumar and Pradhan (2005), '*foreign direct investment (FDI) emerged as the most important source of external resource flows to developing countries over the 1990s and has become a significant part of capital formation*'. Since FDI affects GDP, creating a prejudiced variance for externalities, external influences on economic growth will be controlled for. The measure for FDI is equal to the 'amount of foreign capital liabilities'. This represents the amount of direct investments and foreign equity stock for each year into the regional economy. As described by Kumar and Pradhan (2005), a positive effect on GDP is expected.

4.1.1. Spatial connectivity

The theoretical section indicated the importance of spatial proximity within an agglomeration for externalities and the difference between geographical proximity and spatial connectivity. Where geographical proximity could be implemented effortlessly by measuring the direct distance between two cities, it would not give any insights on any spatial effects on externality interdependencies. Therefore, and according to previous theories, spatial connectivity needs to be implemented.

Guangdong province has experienced major transport developments in the past 20 years. Expressways, inter-city rail networks and specific ferry connections are the most accessible and rational means of transportation within the GPRD (see *Appendix A, fig. A1* and *A2*) - neglecting extraordinary helicopter and aviation transport between cities. The total length of expressways in the GPRD was 1301 km^2 in 2000; 3508 km^2 in 2010; and is around 6150 km^2 nowadays (Hou & Li, 2011). Inter-city train networks have developed in the last 20 years, contributing to faster and more connections, and a higher convenience and frequency. In addition, due to the presence of the Pearl River within the GPRD, ferry transport is a slow but highly adapted means of transportation within and between cities located next to the Pearl River.

Specific data of travel time between each city within the Greater Pearl River Delta for 20 consecutive years is not present in empirical statistics books. Where analysing all direct travel distances for the last 20 years is time-consuming, possibly unachievable and beyond the reach of this study, it will also be impractical as it does not take into account any innovations and improvements for velocity, border controls and infrastructural conveniences (Hou & Li, 2011; Yang, Xia & Zhang, 2018). According to Yang et al., several improvements like faster trains, more capacity and more stations/exits have contributed to a greater level of spatial connectivity between cities within the GPRD over the last 20 years. These adaptations cannot be implemented directly into statistical analysis.

To achieve coherent data about the spatial connectivity between all the cities within the Greater Pearl River Delta, the last official statistics about the direct connectivity will be marginally adapted with Hou and Li's (2011) connectivity matrix. In order to cope with major infrastructural developments which can influence connectivity significantly, major infrastructural developments of the last 20 years will be analysed and data will be adapted if necessary.

Hou and Li's regional connectivity study introduced the previously mentioned connectivity issues, as spatial connectivity '*can be defined as a measure of the potential of being accessed by or to access other locations. To realize this potential, one needs to take into consideration other dimensions, such as availability, affordability, acceptability and accommodation'* (Hou & Li, 2011, p. 1353). These different dimensions are (on average) improving annually due to technological innovations and infrastructural investments. To include these time-dependent variations, Huo and Li aggregated the GPRD into three regions:

- Eastern: HKSAR, Shenzhen, Dongguan, Huizhou;
- Central: Guangzhou;
- Western: MSAR, Zhuhai, Jiangmen, Zhongshan, Foshan, Zhoaqing.

Implying these regional divisions, they empirically derived an average intercity travel time between and within each region for 2000, 2010 and 2020, while taking into account influencing factors (see *table 4.1*). Travel distance within the GPRD decreased around 18% between 2000 and 2010, and 23% between 2010 and 2020. The first reduction is mainly caused by expressway developments, while the latter is mainly a result of gradual investments in the inter-city rail system (see *Appendix A*, *fig. A1* and *A2*).

Year		Eastern *(Index)*	Central *(Index)*	Western *(Index)*	Overall *(Index)*
2000	Eastern	1.12 *(100)*	1.86 *(100)*	2.65 *(100)*	2.12 *(100)*
	Central		1.07 *(100)*	1.70 *(100)*	1.66 *(100)*
	Western			1.54 *(100)*	1.88 *(100)*
2010	Eastern	0.95 *(84.8)*	1.55 *(83.3)*	2.19 *(82.6)*	1.77 *(83.5)*
	Central		0.80 *(74.8)*	1.40 *(82.4)*	1.36 *(81.9)*
	Western			1.25 *(81.2)*	1.54 *(81.9)*
2020	Eastern	0.86 *(76.8)*	1.28 *(68.8)*	1.64 *(61.9)*	1.38 *(65.1)*
	Central		0.70 *(65.4)*	1.18 *(69.4)*	1.15 *(69.3)*
	Western			0.90 *(58.4)*	1.16 *(61.7)*

Table 4.1. *Average intercity travel times (1.00=60 minutes) in 2000, 2010 and 2020 between Eastern, Central and Western regions in GPRD, including time-dependent variations. Adapted from Huo & Li, 2011.*

In order to use these aggregates in panel statistics, an index (100 = average intercity travel time in 2000) is empirically assessed. This allows to gradually include the spatial connectivity improvements in the data.

Using only these aggregate travel distances between the specified regions will generate generalized and non-practical data. Therefore, a base year needs to be included. From this base, the average index differences between the regions will be gradually implemented to involve yearly transport adaptations, (minor) infrastructural improvements and greater local connectedness (Yang et al., 2018). In addition, by including a recent base for the spatial connectivity data, any exposure to empirical estimation errors from the results of Huo and Li (2011) from 2011 till 2020, will significantly reduce.

	HKSAR	**Sh**	**Do**	**Hu**	**Gu**	**MSAR**	**Zhu**	**Ji**	**Zho**	**Fo**	**Zha**
HKSAR	-	60	85	90	120	55	70	125	130	150	135
Sh	60	-	25	30	40	100	100	90	70	90	75
Do	85	25	-	50	25	80	80	70	50	85	75
Hu	90	30	50	-	60	120	120	110	90	130	95
Gu	120	40	25	60	-	60	60	50	30	30	35
MSAR	55	100	80	120	60	-	5	70	30	120	95
Zhu	70	100	80	120	60	5	-	70	30	120	95
Ji	125	90	70	110	50	70	70	-	60	60	85
Zho	130	70	50	90	30	30	30	60	-	90	65
Fo	150	90	85	130	30	120	120	60	90	-	70
Zha	135	75	75	95	35	95	95	85	65	70	-

Note. *Top-left corner: Eastern-Eastern connection*
Bottom-left and top-right corners: Eastern-Western connection
Bottom-right corner: Western-Western connection
Middle rows: Eastern-Central (top) and Western-Central (bottom)

Table 4.2. *Matrix of travel times between cities in the GPRD (minutes, 2015). Adapted from CBRE Research (2016).*

Table 4.2 shows the latest known bilateral travel distances in the Greater Pearl River Delta. This data is derived from CBRE (2016) and indicates all travel times in minutes between the cities. The table indicates once more the importance of a base year instead of taking the aggregate travel distances between regions. For instance, travel time from HKSAR is the shortest to MSAR (55 minutes due to a fast ferry connection) while it is an Eastern-Western connection. Therefore, 2015 bilateral travel distances are used as a base in the data. Using Huo and Li's (2011) indexed regional connectivity matrix, the annual average decrease in travel distance is calculated (see *Appendix B*). This shows that Eastern-Eastern travel times have decreased on average 0.943% annually between 2010 and 2020, and -1.52% annually between 2000 and 2010. By calculating these annual percentual changes the real travel times between each city is calculated. E.g. travel distance between HKSAR and Shenzhen was 60 minutes in 2015. Using the earlier mentioned calculations it indicates a travel time of 72.38 minutes in 2000; 63 minutes in 2010; and 58.3 minutes in 2018. This creates spatial connectivity variables between each city within the GPRD for each year, generating a total of 2090 observations for intercity travel time. Each variable is noted as C(*j*), which indicates the travel time between the specific observation to city *j*. For instance, CHKSAR indicates the connectivity (C) from the selected observation to HKSAR in year *t*.

However, several major infrastructural constructions are planned or have been developed in the past 20 years which can influence these observations significantly. Therefore, these developments need to be analysed. The three key infrastructural developments within the GPRD are the Humen Bridges, Hong Kong-Zhuhai-Macau Bridge and the Shenzhen-Zhongshan Bridge (see *Appendix A, Fig. A2)*.

The first Humen Bridge connects the cities Foshan and Dongguan, covering the total length of the Pearl River of 3.600 meters. It was the first static connection between the Eastern and Western region of the GPRD, reducing travel times between southern located cities significantly (Hou & Li, 2011). The first Humen Bridge, completed in 1997, was quickly insufficient and caused massive traffic problems due to the accelerating economic and population growth. Therefore, the construction of a second Humen Bridge (officially the Nansha Bridge) started in 2013 and was completed halfway 2019. As the first Humen Bridge was opened three years in advance of this dataset, its application is well-defined in the data, and the second Humen Bridge was completed in 2019, it does not influence any observations.

Secondly, the Hong Kong-Zhuhai-Macau Bridge (HZMB) is a 55-kilometre connection between HKSAR and the western cities of Zhuhai and MSAR. The connection consists of three main bridges, covering almost 30 kilometres, a land tunnel next to HKSAR of 1.1km and a 6.7km-long underwater tunnel, four artificial islands, and three boundary crossing facilities (Mainland China, HKSAR and MSAR). This makes the HZMB the longest sea bridge and the longest fixed link in the world (Li, Zhou & Fan, 2015), decreasing the travel time from HKSAR by car to only 30 minutes. Interestingly, the bridge connects three different monetary systems and is a right-driving bridge, even though they drive on the left side of the road in HKSAR and MSAR. The HZMB has been opened on the 24th of October in 2018. However, due to high toll fees, strict immigration processes and unaccustomedness with the new connection, the bridge has experienced mostly tourist transports in the first months (Chen, 2019). Therefore, the HZMB does not alter the observations of 2018 as the opening has not resulted in significant usability in the last two months.

Thirdly, the Shenzhen-Zhongshan Bridge will connect Shenzhen and Zhongshan, reducing travel time to less than 30 minutes. The eight-lane and railway bridge-tunnel connection will become the widest in its sort and is expected to be completed in 2024. Therefore, the development of the Shenzhen-Zhongshan Bridge will not alter any observations.

Analysing the other rail and inter-city highway developments shows that developments have been gradually implemented within the GPRD with one exception. This exception relates to the Guangzhou-Shenzhen-Hong Kong Express Rail Link, also known as the Guangshengang XRL. The first completion of the high speed railway connection, from Guangzhou to Shenzhen, was completed in 2011. This reduced the travel times between the two major cities to 40 minutes. The high speed connection is a major contributor to the connectivity between the two cities since it has a high frequency and functions at the top of its technological capabilities with speeds of around 350km/h. Therefore, an adaptation in the spatial connectivity between Shenzhen and Guangzhou is necessary. The observations between Guangzhou and Shenzhen from 2011 till 2018 will be altered to a static travel distance of 40 minutes. The other section of the Guangshengang XRL towards HKSAR was opened on the 23rd of September 2018, reducing the travel distance between Guangzhou and HKSAR to 50 minutes. However, as travel distances between Shenzhen, Guangzhou and HKSAR have decreased significantly, the data will not be altered. This is due to the same issues present after the operation of the HZMB, namely a significantly low passenger rate of the new-found connection (Yau, 2018). According to Yau, during the first months of operation, the daily passenger rate was almost 40% lower than expected, arguing that passengers need to get familiarized with this new way of transportation. Consequently, alterations of the observations between the spatial connectivity of HKSAR with Shenzhen and Guangzhou for 2018 is not relevant.

4.1.2. Institutional distance

As hypothesized, institutional distance between two cities reduces bilateral externalities as a result of different shared values, cultures and business environments. In the case of the Greater Pearl River Delta, the nine major prefecture-level cities within the Guangdong Province experience the same set of institutions due to similar (centrally governed) institutional development and inter-agglomeration adaptations (Meyer & Diez, 2015). As elaborated in the third section of this study, the Special Administrative Regions of Hong Kong and Macau have a significantly different institutional environment due to its colonial and foreign influences. Therefore, a measure for institutional distance between HKSAR, MSAR and the other nine cities needs to be derived.

Unfortunately, a singular measure for institutional distance is not present and often miscellaneous. This is the case since '*what institutions are and what types of institutions are of interest, differs by level of analysis and by the subject under consideration. As a result, … various measures of institutional distance have been developed'* (Hotho, 2009, p. 5). Firstly, the relevant nature of institutional distance needs to be elaborated.

Three different varieties of institutionalism can be distinguished: new organizational institutionalism, new institutional economics, and comparative historical institutionalism. Hotho (2009) states that the first two theories of institutionalism have a consensus on how to measure or illustrate institutional distance as a quantitative variable. However, in this study, institutional distance needs to be placed inside the field of the third theory of institutionalism: comparative historical institutionalism. This is as comparative historical institutionalism states that institutions are non-static over time and depend on its local environment. This resemblance to path dependency applied in this study is summarized by Hotho (2009):

> *The starting point of the third variety of institutionalism, comparative historical institutionalism, is that the key institutional elements in a society are presumed to develop interdependently over time, resulting in intrinsic differences in economic organization between counties. … . Due to the complex interdependencies between societal features, this strand of institutionalism largely relies on extensive systematic qualitive comparisons to describe such differences between countries.* (p. 6)

So, secondly, to quantify the institutional distance present between the cities of the GPRD, a qualitative set of institutions needs to be compared. Bae and Salomon (2010) elaborate in their review study several dimensions, qualitative operationalization and representative studies in order to capture this mixed set of institutions. They categorize institutions in the following dimensions: political, regulative/legal, cultural and normative distance (Bae & Salomon, 2010). Where international economists often focus on formal institutions (political and regulative/legal institutions) and social studies focus on normative and cultural dimensions of institutions, this study aims at including a mix of several dimensions of institutions. This is to favour the historical embeddedness of path dependency, which is present in both formal institutions and in culturally adapted virtues. Therefore, this study applies a diversified set of several institutional dimensions which first need to be elaborated.

Political institutions refer to government structures and policies. High political institutional distance increases transaction costs as it is more challenging to compete and conduct business effectively. For instance, the security of property rights, contract enforcements and the political system are more at risk in the case of uncertain political institutions. Regulative or legal institutions relate to the enforcement of regulations and its local enactments. As stated by Bae and Salomon (2010), regulative and political institutions can be intertwined in the case of minor transparency in the bureaucratic and legislative decision making. This means that the discrepancy between legislative institutions who control the security of property rights, and political institutions that implement the rules, is rather vague. This relationship varies amongst countries and needs to be carefully analysed according to its context. Therefore, in this study including the Greater Pearl River Delta, political and regulative institutions will be considered as one intertwined dimension.

Economic institutions refer to the economic and market structure, and its related implementations. A greater economic resemblance encourages economic exchange and attractiveness to conduct business. However, the causality of the development of economic institutions is often debated and ambiguous (Bae & Salomon, 2010; Krugman & Obstfeld, 2006). Bae and Salomon describe that the economic environment is often initiated by cultural and political principles. They state '*that patterns of exchange, economic structure, and so on may be viewed as the result of firms and markets reacting to political, regulatory, and cultural institutions'* (Bae & Salomon, 2010, pp. 332-333). Another debate relates to the role of geographical proximity and historical connections on the presence of economic institutions (Krugman & Obstfeld, 2006). Due to these considerations and possible multicollinearity issues of related variables in this study, economic institutions will not be included in the analysis. Since the other institutional dimensions mentioned by Bae and Salomon (2010) and the role of geographical proximity and path dependency are included in this study, relevant institutional data will not be excluded.

Regarding informal dimensions, cultural distance relates to different values and beliefs within the society. Similar beliefs and cultural values enhance means of communication and coordination between regions. If a region experiences a high level of cultural distance, mutual understandings and shared trust decreases, reducing the overall level of operational effectiveness (Hotho, 2009). Cultural distance needs to be carefully differentiated from normative distance. Normative institutions contribute to and determine (social) '*rules, including routines, roles, procedures, conventions and codes,*' which *'provide the framework for ordering action and explain both the organizational structures and the actions of individuals within those structures'* (Alexander, 2012, p. 796). Normative institutions thus govern actions and determine local societal structures, contributing or harming the operational effectiveness. These institutions refer more to the managerial attitudes compared to an entire society's culture and beliefs.

Thus, in practicalizing a measure for institutional distance from the perspective of comparative historical institutionalism, three dimensions need to be included: political/regulative (legal) institutions, which monitor and enforce legal rules and regulations; cultural institutions, which illustrate shared beliefs and values within the society; and normative institutions, which determine social norms and conventions based on local standards. Finding appropriate empirical measures is essential to capture each proposed dimension. Based on studies related to institutionalism, several measures are selected that describe the legal, cultural and normative dimension of a local environment. *Table 4.3* lists these indicators, including the origin and representative studies.

Dimension	*Operationalization based on*	*Representative study*	*Indicators[1]*
Legal: Political and regulative	Global Competitiveness Report (World Economic Forum)	Chao and Kumar (2010), Gaur and Lu (2007), Xu, Pan and Beamish (2004)	Intellectual property protection (1.01) Judiciary system efficiency (1.06) Settlement of disputes (1.10) Political transparency (1.12) Effectiveness of police force (1.16) Investment protection (1.21) Effectiveness of anti-trust laws (6.03)
	Economic Freedom Index (The Heritage Foundation)	Ionascu, Meyer and Estrin (2004)	Government integrity
Cultural	Hofstede Index	Cho and Padmanabhan (2005), Hotho (2009), Ionascu et al. (2004)	Power distance Individualism / collectivism Masculinity Uncertainty avoidance Long-term orientation Indulgence / restraint
Normative	Global Competitiveness Report (World Economic Forum)	Xu et al. (2004)	Efficacy of corporate boards (1.19) Customer orientation (6.15) Labour-employer relation (7.01) Efficient use talent (7.02) Performance-related pay (7.06) Reliance on prof. management (7.07)

Note. [1] Corresponding items in Global Competitiveness Report are in parenthesis.

Table 4.3. *Measurement items of legal institutions (indicator 1 to 8), cultural institutions (9 to 14) and normative institutions (15 to 20).*

Indicators like intellectual property protection and judiciary system efficiency refer to the implementation of local rules and regulations in the economy, while political transparency and investment protection relate to the legitimacy of doing business (Gaur & Lu, 2007). The data sources, the Global Competitiveness Report and Economic Freedom Index are based on survey responses and are published annually by respectively the World Economic Forum (2019) and The Heritage Foundation (2019). These sources are often used for institutional analyses due to its standardization and global reach (Hotho, 2009). Regrettably, the applied methodology of the Global Competitiveness Report has been altered since 2007. This generates less specified and non-comparable data for the periods 2000 till 2006, and 2007 till 2018. In order to eliminate the risk of biased data between these periods and to comply with the appropriate measures within comparative historical institutionalism, its indicators will be utilized for observations succeeding 2007.

Regarding the cultural dimension, institutional distance will be measured applying Hofstede's cultural dimensions: power distance, individualism/collectivism, uncertainty avoidance, masculinity, long-term orientation, and indulgence/restraint (Hofstede, 2001; Hotho, 2009). These six dimensions are widely applied as cultural and societal indicators for national economic and cultural differences, as well as a cultural measure for cultural institutional distance (Ionascu et al., 2004). However, as the cultural dimensions of Hofstede are measured nationally, no specific measures for cultural distance between the cities in the Guangdong province are present. Nevertheless, Hofstede's cultural dimensions are present for Hong Kong and Macau SAR, which makes it a proper comparable measure amongst the cities (Ionascu et al., 2004). In addition, as cultural distance is a subset of institutional distance in general in this study, a too anticipated effect of similar cultural distances within the Guangdong province is ruled out. Since culture is seen as a rooted and static local occurrence, the measures are analogous during the entire phase of analysis (Hofstede, 2001).

Normative institutional distance is measured by using several measures from the Global Competitiveness Report. Whereas normative institutional distance refers to managerial varieties, the freedom and degree for managerial positions (relate to the efficacy of corporate boards, reliance on professional management and efficient use of talent) and the relative position and customs within firms (customer orientation, labour-employer relations and performance-related payments) are correct measures for normative institutions (Xu et al., 2004).

Unfortunately, the Global Competitiveness Report (World Economic Forum, 2019) does not include measures and indicators related to Macau SAR. Other studies focussing on the institutional structure of MSAR have faced similar lacks of data and often argued that the institutional framework of MSAR is identical to the institutional framework of HKSAR. Wong and Kwong (2020) state that *'under the OCTS with the principles of 'Hong Kong/Macau people ruling Hong Kong/Macau' and 'a high degree of autonomy,' the institutional arrangements of the two SARS are virtually identical.'* (p. 47). In addition, Sadowski (2016) highlights the legal institutional similarity between both SARs created by the OCTS. He states that '*the legal bilinguality is one of the reasons why Hong Kong and Macau are thought to be 'extraordinary' .... However, what is particularly striking in the two SARs, is the level of local autonomy granted to them in the Basic Laws'* (2016, p.12). Where it decreases the credibility of the specific measure for normative- and legal distance for the Special Administrative Region of Macau marginally, the overall effect and increase of reliability for institutional distance by adapting similar values of HKSAR are acknowledged to expressively positively compensate the aforementioned.

The final step in generating a measure of institutional distance is the calculation of a single empirical measure for each distance. The singular values of the indicators will be annually (except for cultural distance) combined into a composite index without correcting for differences in variance. As mentioned by Hotho (2009), '*correcting for variance imposes certain weights on the indicators included in a composite index. When the original scores are scaled similarly, as is the case in the Global Competitiveness Report, correcting for variance would inflate relatively small institutional differences'* (p. 29). As this study also applies measures from the Global Competitiveness Report and other broadly applied sources, correcting for variance might only result in biased, distorted or inflated effects.

Therefore, each original measure will be scaled similarly and applied to one of the following equations:

$$Legal\ ID_{jk} = \sum_{i=1}^{8}(I_{lj} - I_{lk}) / 8 \tag{4.1}$$

$$Cultural\ ID_{jk} = \sum_{i=1}^{6}(I_{cj} - I_{ck}) / 6 \tag{4.2}$$

$$Normative\ ID_{jk} = \sum_{i=1}^{6}(I_{nj} - I_{nk}) / 6 \tag{4.3}$$

These equations specify the measure for institutional distance, consisting of the composite index (I) for legal (l), cultural (c), and normative (n) institutions, between country/region j and k. The results of each equation generate an index measure for legal, cultural and normative institutions for each year and region. In order to avoid distorting or biased indicators and thus ensure the validity of the measures, each distance index will be tested for multicollinearity. Cronbach's alpha (α) is measured for each institutional dimension to verify the conceptual relatedness of each indicator. A correct alpha confirms an internal consistency of the composite index, and thus a high uniformity for each dimension of institutions. According to Taber (2018), α needs to be greater than 0.75 in order to ensure high reliability. Nevertheless, an alpha far greater than 0.95 might indicate multicollinearity between several indicators.

The composite index of the eight indicators for legal institutional distance have an α of 0.9623, which thus indicates a certain degree of redundancy (Taber, 2018). In order to eliminate this, the correlation matrix of the variables within the composite index needs to be analysed. This matrix shows that the indicators 'effectiveness of police force' and 'judiciary system efficiency' are the strongest correlated with the other measures. Where 'judiciary system efficiency' might be correlated due to political and regulative connections, the strong correlations for 'effectiveness of police force' have no theoretical or logical rationales. Nevertheless, excluding these two indicators with the highest correlation coefficients with the other indicators for the composite index, shows an α of 0.9284, which indicators no multicollinearity. The composite index for cultural institutional distance has an α of 0.8434, and thus needs no adaptations. The six indicators for normative institutional distance have an α of 0.9763 and thus also faces a certain degree of redundancy. Using the correlation matrix, 'customer orientation' and 'performance-related pay' shows the greatest correlations with other indicators. Leaving these two measures out of the composite generates an admissible α of 0.957.

Hence, the composite indexes of legal and normative institutional distance will consist of respectively six and four indicators to eliminate redundancy issues. This generates empirical measures to indicate institutional distance for each dimension. As the measure for the Guangdong province applies to all nine prefecture-level cities of Guangdong province, this results in a steady amount of observations for legal, normative and cultural institutions (see *table 4.4*).

Nonetheless, in order to avoid empirical issues due to a lack of sufficient observations, data for legal and normative institutional distance for the period of 2000 till 2006 needs to be included. Referring to legislative institutional distance, Guangdong and HKSAR show a variation of respectively 6,657 and 3,149. Where the disparity of Guangdong is greater than HKSAR, the legal index for Guangdong is constant over time (except for the observation for Guangdong in 2007). Theoretically, there are no justifications for the relatively larger change in legal institutional distance in 2007, as well as any other developments between 2006 and 2000. Therefore, it is theoretically grounded to assume that the indicators for legal institutional distance are constant in this period. The average of the empirically derived indicators is calculated and applied as the measure for the years 2000 till 2006. This results in a value of 56,140. For HKSAR, the gap is smaller but the values are diminishing gradually. As stated by Bie et al. (2015) and Vogel et al. (2010), this might be the case due to the colonial past of HKSAR. The autonomous region faced different legislative and political institutions and is gradually adapting to the legal standards of the PRC. Therefore, in order to align the missing values for the legal dimension between 2000 and 2006, values will be calculated using an average percentual alteration. Between 2007 and 2018, legal indicators decreased by 2,537%, which is 0,231% annually. So, theoretically, starting from 2007 values are annually calculated using a gradual change of 0,231%.

Regarding the normative dimension, there is a relatively stable pattern of index values. As described by Huesmann and Guerra (1997), factors responsible for normative behaviour are an interaction of social behaviour and not very flexible. Nevertheless, the indexes for the normative dimensions have a difference of 3.502 (Guangdong) and 6.518 (HKSAR), indicating some degree of variation over time. Firstly, any reasons behind these gaps need to be derived. The indicators efficacy of corporate boards, efficient use of talent, and reliance on professional management, all have lower values in the years 2009 till 2014. Only the indicator for labour-employer relations stays equal. However, values of 2008 and 2007 are similar to values after 2014, leaving room for speculating thoughts. A broad literature examination results in only one valid clarification for these results, namely the local impact of the financial crisis in HKSAR. As stated by Lee et al. (2010), *'economic contraction triggered by a global financial crisis was associated with a significant increase in the risk of depression in the Hong Kong population'* (p.125). In combination with HKSAR functioning as the financial centre of the GPRD, the financial crisis could have influenced normative behaviour within the city and thus alter normative indicators temporarily. Therefore, the average values of the years 2007, 2008, and 2013 till 2018 is calculated and determined for the missing values. Values related to Guangdong surprisingly show a minor increase in values in the years 2008 till 2011, one year prior to the shift of HKSAR's values. Theoretical explanations for this variance are absent, except for the financial crisis. Therefore, and in line with the calculations of the missing values for HKSAR, the average of the values of 2007, and 2012 till 2018 is taken. Values for 2008 are excluded (and values for 2012 included) since those values are considered as outliers. This generates two values for the missing values for Guangdong and HKSAR: respectively 64.697 and 80.770. All values for the legal, normative and cultural dimensions of institutional indexes are presented in *table 4.4*.

	Legal dimension		**Normative dimension**		**Cultural dimension**[1]		
Year	***Guangdong***[2]	***HKSAR***[3]	***Guangdong***[2]	***HKSAR***[3]	***Guangdong***[2]	***HKSAR***	***MSAR***
2018	57.071	80.933	63.458	83.135	51.178	42.820	45.667
2017	56.470	80.664	63.494	81.873			
2016	54.948	80.097	65.024	80.174			
2015	55.095	79.930	64.533	80.174			
2014	55.846	81.442	65.489	78.865			
2013	56.121	82.195	65.127	79.052			
2012	56.495	80.811	65.167	76.673			
2011	58.061	81.022	67.413	76.617			
2010	58.237	82.625	67.226	78.212			
2009	58.011	81.402	67.850	77.692			
2008	55.745	82.437	68.786	80.701			
2007	51.580	83.080	65.284	82.190			
2000-2006	56.140	84.430[4]	64.697	80.770			

Note. [1] *Annual static measure for cultural distance.*
[2] *Represents the nine cities of the GPRD within the Guangdong province: Guangzhou, Shenzhen, Dongguan, Huizhou, Zhuhai, Jiangmen, Zhongshan, Foshan and Zhaoqing*
[3] *Similar values for MSAR*
[4] *Value for year 2000 is given*

Table 4.4. *Composite indexes for legal, normative and cultural institutional distance for Guangdong, HKSAR and MSAR. All measures scaled similarly (1-100).*

4.1.3. Industry similarity

As hypothesized in the theoretical section, greater industry similarities between two cities within an agglomeration result in localization economies. Consequently, this effects the microlevel interrelatedness, generating a greater presence of externality interdependencies.

In order to measure industry similarity, the economies of all cities need to be contrasted and compared. Since a specific generalized measure for each industry or sector is missing in the database of the GSB, the percentage of GDP of each industry is used to calculate industry similarities. The datasets of the GSB, the Census and Statistics Department (2019) of the Government of Hong Kong and the Statistics and Census Department (2019) of Macau SAR, divide industries into four main categories: primary industry; manufacturing; construction; and tertiary (service) industry. Each database has comparable industry specifications except for waste management for Hong Kong SAR. Interestingly, waste management is included in the category 'manufacturing' and mentioned in abundance. This can be related to the major waste problems that the city has been facing for years (Lu & Tam, 2013). Nevertheless, waste management does not generate a significant output that it effects or biases the industry measures.

In order to measure industry sector similarity, Marll's (2008) measure is adapted. His measure captures the degree of industry structure dissimilarity (deviation) between city *i* and city *j*. For each industry category the percentages of GDP are taken, and for two given cities, subtracted from each other. For instance, Foshan's and Huizhou's economies were in 2018 respectively 54.4% and 49.5% (of GDP) dependent on the manufacturing industry. The difference in manufacturing between both economies is thus 4.9%. This value thus represents the deviation in industry similarities (Marll, 2008). Since these variables measure a degree of industry dissimilarity, it is expected that the effect of these industry dissimilarity measures on externality interdependencies becomes negative.

Examining the data shows the global trend towards a tertiary industry, reducing the role of the primary industry for economic growth between 2000 and 2018 (*Appendix C)*. This transformation is also represented by each categories' mean and standard deviation which are respectively 4.65 and 6.48 (primary); 2.92 and 1.22 (construction); 39.90 and 18.04 (manufacturing); and 52.52 and 20.21 (tertiary sector). Where the latter two industry categories contribute economically the most to the agglomeration, the former two categories have a minimal impact and almost no disparities between cities due to the low standard deviations. This will reduce the validity due to an increased focus on minor variances. In order to avoid biases of insignificant indicators, 'primary' and 'construction' industry will be excluded from the analysis. In addition, the relative impact of both industry categories is expected to be minimal due to its minimal share of local GDP and the low-technological intensity, which does not influence the presence of externality interdependencies significantly (Artal-Tur et al., 2012).

This generates a total of 22 variables, indicating bilateral industry deviation for the 'manufacturing' and 'tertiary' industry, with 209 observations each. The values of 2010 for both manufacturing and tertiary industry are represented in *table 4.5.* Each variable is named according to the industry category and city *j*. For instance, TerJi indicates the industry disparity for the tertiary industry between the observation and Jiangmen in year *t*. Since this measure for industry deviation consists of two separate industries, they need to be measured independently in order to eliminate correlation issues and conflicting internal measures. Therefore, the tertiary and manufacturing sector need to be two separate indicators, which will be discussed in the estimation section. Furthermore, as both categories are substitutable a high degree of (negative) covariance is expected when measuring bilaterally.

		Manufacturing industry										
		HKSAR	**MSAR**	**Gu**	**Sh**	**Zhu**	**Fo**	**Hu**	**Do**	**Zho**	**Ji**	**Zha**
Tertiary industry	**HKSAR**	-	0.1	30.1	40.4	47.4	56.7	51.7	45.1	51.4	49.3	34.1
	MSAR	2.1	-	30	40.3	47.3	56.6	51.6	45	51.3	49.2	34
	Gu	32	34.1	-	10.3	17.3	26.6	21.6	15	21.3	19.2	4
	Sh	40.3	42.4	8.3	-	7	16.3	11.3	4.7	11	8.9	6.3
	Zhu	50.5	52.6	18.5	10.2	-	9.3	4.3	2.3	4	1.9	13.3
	Fo	57.6	59.7	25.6	17.3	7.1	-	5	11.6	5.3	7.4	22.6
	Hu	57.9	60	25.9	17.6	7.4	0.3	-	6.6	0.3	2.4	17.6
	Do	44.3	46.4	12.3	4	6.2	13.3	13.6	-	6.3	4.2	11
	Zho	53.8	55.9	21.8	13.5	3.3	3.8	4.1	9.5	-	2.1	17.3
	Ji	56	58.1	24	15.7	5.5	1.6	1.9	11.7	2.2	-	15.2
	Zha	52.6	54.7	20.6	12.3	2.1	5	5.3	8.3	1.2	3.4	-

Table 4.5. *Values for industry disparities for 'manufacturing' and 'tertiary' for 2010.*

4.2. SEM: externality interdependencies framework

The following section focusses on the estimation techniques and model building for this study. Based on the theoretical framework and methodology, the applied statistical technique is substantiated and validated.

As formulated earlier, the theoretical framework for externality interdependencies is captured in the simultaneous equation model (*2.4* and *2.5)*:

$$y_{i,t+1} = \sum_{j=1}^{n} X_{ij,t} + \beta_2 C_{i,t} + u_{i,t} \tag{2.4}$$

$$\mathrm{X}_{\mathrm{ji,t}} = \mathrm{y}_{\mathrm{j,t}} \left(\mathrm{S}_{\mathrm{ji}}, \mathrm{ID}_{\mathrm{ji}}, \mathrm{I}_{\mathrm{ji}}, \mathrm{C}_{\mathrm{j}}\right) \qquad (2.5).$$

Since the concept of externalities includes both Jacobs' and MAR externalities, which are generated through several means and channels, a completely conclusive dataset is an unrealistic assumption. Therefore, a fixed effects model is implemented as certain variables are rationally omitted (Williams, 2018). In addition, a fixed effects model is more applicable if variables have a higher degree of variability. As shown in *table 4.7¸* within-subject variability is present and thus acknowledges the adaptation of a fixed effects approach (Williams, 2018). As elaborated in the theoretical section, in order to include the path dependent approach in this study, agglomerations are considered as non-static. A dynamic fixed effects model allows to include economic growth as a lead variable, to include a time dimension on the interaction between externality interdependencies and economic growth (De Groot et al., 2009).

In order to statistically test the simultaneous equation model (*2.4* and *2.5*), a multivariate statistical method needs to be considered (Harlow, 2014). Several multivariate analyses are available. Nevertheless, in order to establish a framework including a path dependent approach in this study, a structural equation model (SEM) is applied. SEMs are often build upon theoretical constructs and the relationships between these constructs. Often, SEMs are named covariance structure models since it's statistical measurements are dependent on the covariances between observed variables (Harlow, 2014).

A structural equation model is selected for this study due to its comprehensive possibilities and implementations since it '*provides a very general and convenient framework for statistical analysis that includes several traditional multivariate procedures'* (Hox & Bechger, 1998, p. 354). The main reasons for implementing a SEM in this study are as follows.

Firstly, a SEM includes both manifest and latent variables. Observed variables are variables that are measured and of which data is available and collected. The opposite of an observed variable is a latent variable, often also called a factor or construct (Burnette & Williams, 2005). Latent variables are unobserved variables. As stated by Burnette and Williams (2005), '*a latent variable is a variable that cannot be directly observed and must be inferred from measured variables'* (p. 146). The measured variables (observed variables) are thus indicators for the often hard to measure latent variables. Thus, a latent variable captures the relation of a concept, which is measured by observed variables, within a structural model. Often within agglomeration studies, variables or developments are unobservable or partially explained by observations. For instance, institutional distance is a composite of the indicators for normative, cultural and legal institutional distance, which is unmeasurable as one variable. Latent variables provide the opportunity to practice variables that cannot be measured using other multivariate analyses. Within SEM, it is assumed that latent variables have a correlation effect on each other, and thus contribute complementary to the presence of externality interdependencies.

Additionally, the theoretical framework states that path dependency generates a higher degree of economic integration, effecting externalities within the agglomeration (Press, 2006). Thus, externalities are effected by simultaneous developments within the agglomeration, as well as interactions between these developments (Sunley, 2006). According to Harlow (2014), structural equation models are well fitted for analysing interaction effects and simultaneous equation models (see *2.4* and *2.5*). She states that '*SEM allows modelling a nonlinear relationship between the latent variables, for example, quadratic and interaction effects amongst the latent variables'* (p. 30). This allows testing the relationships between determined variables on externalities, and their interactions within the agglomeration. Thus, applying a structural equation model allows to measure unobservable variables, interactions between those variables, and provides a common and multi applicable statistical model for agglomeration studies.

Furthermore, capturing and measuring the specific outgoing and incoming externalities from one city within the agglomeration is the objective. Multigroup SEM (MSEM) allows measuring group-specific externality interdependencies to be measured, either outward- or inward-bound. These are also illustrated by *eq. 2.6a* and *2.6b*, which analyse the specific impact of a city within the agglomeration. A MSEM enables group comparison, which examines the influence of different degrees of the pre-specified concepts as well as the role of different degrees of economic growth from the city of study. Nevertheless, multigroup SEM (MSEM) needs to be carefully approached. Qureshi and Compeau (2009) compared the results of MSEM depending on several data specifications, as well as the data formation. Therefore, MSEM requires a careful approach in order to avoid misinterpretations and biased results.

Since the theoretical framework and prior studies provide consistent and broadly analysed relations, Confirmatory Factor Analysis (CFA) is applied. Related to CFA, Byrne (2010) states that:

> *In contrast to EFA* [Explanatory Factor Analysis]*, confirmatory factor analysis (CFA) is appropriately used when the researcher has some knowledge of the underlying latent variable structure. Based on knowledge of the theory, empirical research, or both, he or she postulates relations between the observed measures and the underlying factors a priori and then tests this hypothesized structure statistically.* (p. 6)

Within CFA, relations and its correlated factors can be analysed based on variances of the construct of the SEM and different factor loadings. Therefore, in line with the path dependent approach, results of any developments or related variables can be tested for any variances in factor loadings.

To formulate the MSEM and CFA, the relations of the variables, and the measurement and structural model need to be described. As earlier elaborated, exogeneous and endogenous variables firstly need to be differentiated. It is noteworthy that within SEM a latent variable is considered as exogenous in the measurement model and its indicators as endogenous (compared to exogenous in the theoretical framework). This is since the relations between the variables are causal effects and it is assumed that the observed variables reflect the latent variable. To capture the exact effect and significance of spatial connectivity, institutional distance, and industry similarity (consisting of two industries) on externality interdependencies, four latent variables are constructed. In line with the theory and econometric model, these composed latent variables then have a causal effect on the presence of externality interdependencies for economic growth. Thus, a fifth latent variable is constructed to measure the total effect of externality interdependencies. In addition, within MSEMs each endogenous variable has an error term which reflects variances not incorporated in the model. These specified conditions of each variable are summarized in *table 4.6*.

	Endogenous				**Exogenous**
Observed	GDP_{t+1};	CJi;	ManZhu;	TerGu;	Human Capital;
	CHKSAR;	CZha;	ManFo;	TerSh;	Wages;
	CMSAR;	Legal ID;	ManHu;	TerZhu;	Population;
	CGu;	Normative ID;	ManDo;	TerFo;	Foreign Direct Investments.
	CSh;	Cultural ID;	ManZho;	TerHu;	
	CZhu;	ManHKSAR;	ManJi; ManZha;	TerDo;	
	CFo;	ManMSAR;	TerHKSAR;	TerZho;	
	CHu;	ManGu;	TerMSAR;	TerJi;	
	CDo;	ManSh;		TerZha.	
	CZho;				
Latent (unobserved)	ExterInter (*Externality Interdependencies*)				SC (*Spatial Connectivity*); ID (*Institutional Distance*); IndMan (*Industry Deviation Manufact*); IndTer (*Industry Deviation Tertiary);* Measurement errors[1].

Note. [1] *Refers to all errors for all endogenous variables (38 in total).*

Table 4.6*. Conditions of all variables for structural equation modelling.*

Secondly, based on these concepts and the implementation of multigroup analysis within the SEM, the measurement model can be represented as a path diagram. To visually interpret the model, the theorized relations first need to be expressed. Within SEM, causal effects between variables in SEM are indicated by arrows, also called 'paths' (Hox & Bechger, 1998). They are represented by single headed arrows. Double headed arrows indicate covariances, which eliminates a causal relation but states that two variables covary (Hox & Bechger, 1998). Covariances are estimated pre- and post-estimation; can be depended on bilateral correlations between observed variables or covariances in the estimated error terms. Furthermore, observed variables are represented by a square box and latent variables as an eclipse.

In line with the theorized relations and specified variable conditions, a path diagram can be generated. This path diagram is represented in *fig. 4.1*. The latent variables *SC, ID, IndMan* and *IndTer*, which are determined by the relevant indicators, are concurrently effecting externality interdependencies (represented by latent variable *ExterInter*). Automatically, the path diagram includes 'fixed' error terms for each endogenous variable, as well as one factor loading between an indicator and the corresponding latent variable. Highlighted in red are the theorized covariances between *SC* and *ID*, as well as between *IndMan* and *IndTer*. As mentioned before, a SEM assumes that exogenous latent variables in the measurement model are correlated. These covariances are incorporated in *fig. 4.1*. Identification specifications of the model are elaborated broadly in the analysis section.

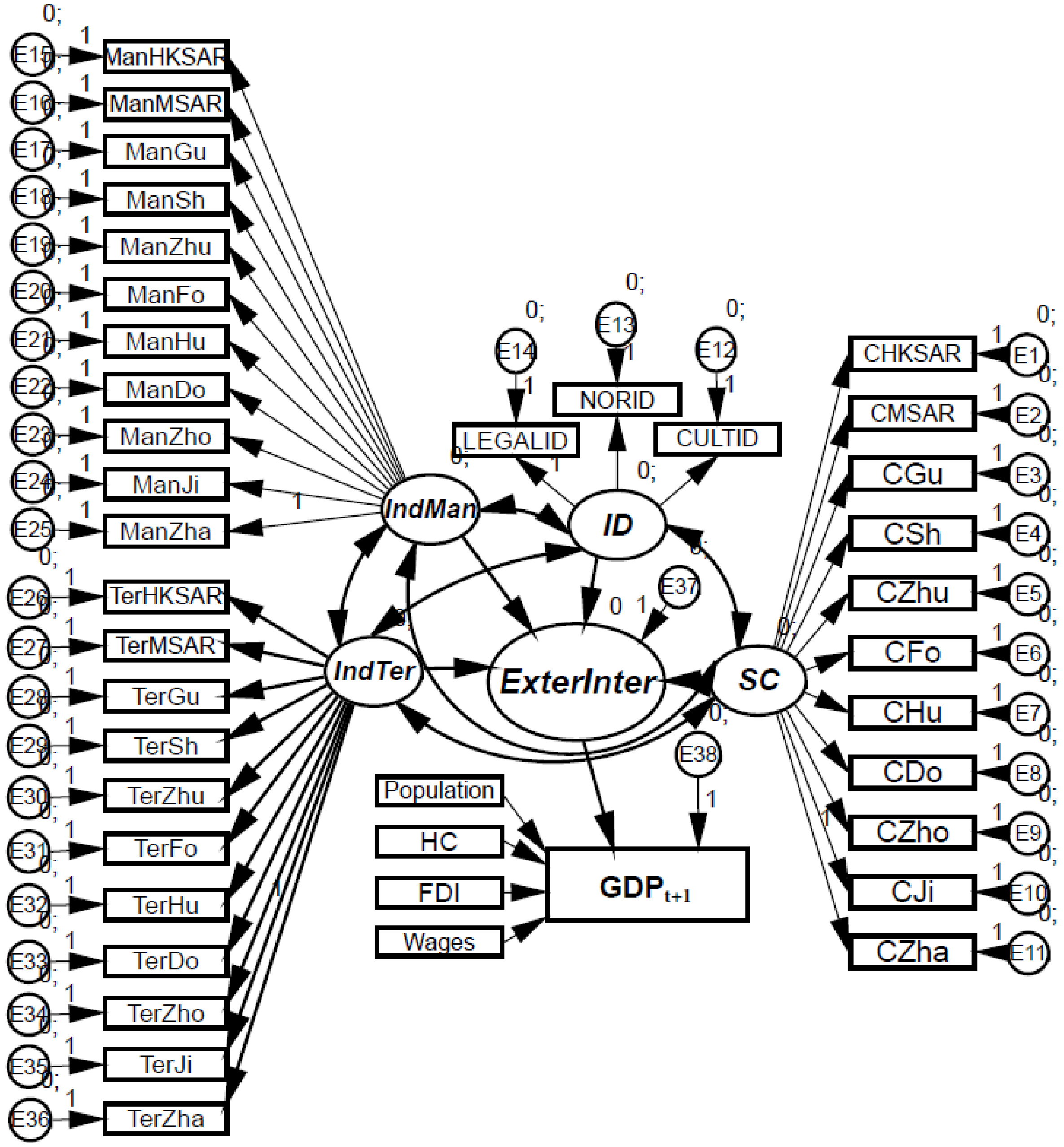

Figure 4.1. *Path diagram of the structural equation model for externality interdependencies.*

4.3. Descriptive statistics and data analysis

The following section focusses on the descriptive statistics and the data analysis. Based on the descriptive statistics presented in *table 4.7* the data will be analysed and adapted if necessary. Furthermore, to eliminate biased results of the SEM several preliminary analytical tests are performed, on top of recommendations for the post-test procedures.

Firstly, the derived statistics presented in *table 4.7* are examined and described. Regarding the control variables, a significant increase is present in all variables. Gross Domestic Product (as presented in *figure 3.2*) has increased drastically with an average annual growth value of 12.375%. The highest average growth rates can be accounted to Shenzhen, Guangzhou and Foshan, while Macau SAR has the highest growth rate in general (30.9% annual growth in 2010). On average, a city within the Greater Pearl River Delta has a population of 5.4 million inhabitants. Guangzhou has been the biggest city by means of inhabitants over the last two centuries, while Macau SAR has always been the smallest.

After applying annual average exchange rates between the Macanese Pataca, Hong Kong Dollar and Chinese Yuan, it becomes clear that the wages in Hong Kong SAR have been the highest for the last two centuries, with Macau SAR coming in second. Whether labour regulations or minimum wage restrictions are a determinant for the relatively high wages of the Special Administrative Regions (compared to the cities in the Guangdong province) is unknown. Nevertheless, the wage gaps between the SARs and nine cities of the Guangdong province have been decreasing over years. Especially Shenzhen and Guangzhou have faced radical wage growth over the last 18 years, with average annual wage growth rates of respectively 9.28% and 10.36% (compared to annual wage growth rates for HKSAR (2.78%) and MSAR (5.97%)). These quantitative visualizations of the great economic growth that has occurred in the nine prefecture-level cities of the Guangdong province, and especially in Shenzhen and Guangzhou, is also present when considering the new product output values (*Human Capital*) as well as the foreign capital utilized (*FDI*) by each city. Human capital has increased drastically within the GPRD. Not only in the core cities of the agglomeration there have been annual growth rates of around 20%, but also the (relatively) small cities have generated tremendous human capital. Theoretically, these growth rates can be accounted to the increased manufacturing productivity in the smaller, labour intensive, prefecture-cities, while technological innovation could be regarded as the main determinant for human capital growth for the largest cities in the GPRD. Regarding the foreign direct investments, the data shows that the more technologically intensive cities Shenzhen and Guangzhou have experienced significantly more FDI, compared to the manufacturing intensive cities within the GPRD. Whether these specific industry characteristics are also the main determinants for these differences in foreign direct investments is unknown.

Regarding the spatial connectivity between each city within the Greater Pearl River Delta, the average transportation time in minutes has decreased on average with 35% between 2000 and 2019. For instance, the longest travel time of 198 minutes in 2000 was between Foshan and HKSAR. In 2018, this connection still accounts as the longest (in minutes) with a travel time of 138 minutes. Nevertheless, it becomes clear that on average the connectivity between all cities within the agglomeration has decreased significantly.

The data on normative, legislative and cultural institutional distance show interesting results. As elaborated before normative institutional distance remains quite stable over the two centuries, while the measure for legislative institutional distance for HKSAR and MSAR is converging towards the values of the Guangdong province. This could be explained by the increased influence of the CCP over the former colonies due to the implementation of OCTS. As stated by Vogel et al. (2010), legislative institutions are less uncomplicated to adapt to new standards and norms compared to more cognitive bounded institutions (normative and cultural). This is also confirmed when considering the applied measurements represented in *table 4.3*. The indicators related to normative institutional distance are more embedded in personal behaviour and organizational performance, while the indicators for legislative institutional distance can be considered as more practical and commutative.

Regarding the industry deviation, the data indicates that HKSAR and MSAR have a distinctive industry structure compared to the cities of the Guangdong province (except for Guangzhou). Between 2000 and 2018, the economies of the two former colonies have had the highest percentage of industrial activity in the tertiary sector within the agglomeration. Nevertheless, the gap has been diminishing over the last couple of years. Whereas all cities have experienced some sort of change of industry priority from the manufacturing to the tertiary sector, this transition is mostly present in Guangzhou. In industrial terms, Guangzhou can be positioned between the manufacturing cities like Jiangmen and Huizhou and the two SARs. It can be assumed that this industrial structure is a result of the previously elaborated (economic) position of Guangzhou within the GPRD, which states that Guangzhou was the political, cultural, and transportation hub since the 1980s.

Secondly, the data needs to be assessed for normality. Structural equation modelling estimates parameters which are derived using maximum likelihood (ML) estimations. As elaborated by Yuan and Bentler (2007), *'parameter estimates are obtained by maximizing the likelihood function derived from the multivariate normal distribution. Standard errors (SE's) of the maximum likelihood estimators (MLE) are based on the covariance matrix that is obtained by inverting the associated information matrix'* (p. 368). In order to apply maximum with for estimating SEMs, the assumption of multivariate normality needs to be met. Stated by Allison (2003), '*when this assumption is met, ML produces parameter estimates that have optimal large-sample properties: consistency, asymptotic efficiency, and asymptotic normality'* (p. 548). Thus, to generate unbiased ML estimations, non-normal distributed data needs to be improved. Preliminary, the data is inspected for multivariate normality using numerical methods, while post-estimation model inferences related to robust procedures can be implemented, if appropriate. To ensure the reliability of the SEM, variables exceeding critical terms related to normal distributions need to be transformed, revised or deleted (Harlow, 2014). The normality of the distributions is tested preliminary as well as post-test to ensure unbiased ML estimations. Preliminary, the data is tested using two numerical methods. Firstly, the presence of outliers is tested, followed by relevant by analysing skewness and kurtosis test statistics.

A graphical review of all observations shows no significant data issues or outliers, except for FDI and GDP. Although one solution includes dropping the observations including these uncertain observations, it is not the preferred option. This is the case since '*pursuing a multivariate normal distribution by deleting observations should be balanced against loss of model power in the interpretation of the results'* (Gao et al., 2008, p. 1). These observations are main contributors to the visualization of the rapid economic growth in the GPRD, next to the relatively small sample size. Thus, the relevant observations are not eliminated.

The descriptive statistics presented in *table 4.7* comprises the skewness and kurtosis values for all variables. Both measures are indicators for univariate normality. Skewness refers to the asymmetry of the normal distribution of a variable. Non-normal distributed variables with skewness values exceeding 1.0 greatly are considered to be positively skewed. Conversely, when skewness values are far lower than -1.0, data is considered to be negatively skewed. A non-normal distributed variable with a high value for kurtosis indicates that the variable has extreme values in either or both tails of the distribution. Kurtosis values lesser than -2.0 or greater than 2.0 are considered to indicate a non-normal distribution (Harlow, 2014). However, Harlow takes into account values related to the excess kurtosis, while STATA measures values for kurtosis. The variance of excess kurtosis compared to STATA's kurtosis is a correction term of -3, with the intention of simplifying the interpretation of the given values. Thus, in order to examine the data for kurtosis the values of Harlow need to be corrected with the correction term, generating acceptable values for kurtosis between 1.0 and 5.0.

The variables GDP and Human Capital have values exceeding the critical values for skewness, while FDI distribution exceeds both skewness and kurtosis. Therefore, all three variables are transformed using the common log transformation to generate a statistically normal distributed variable. This reduces the values of all three variables within acceptable terms for both measures, as well as any biased results due to the previously mentioned presence of outliers for GDP and FDI. The variable Wages is slightly skewed but is not expected to cause major issues due to the satisfactory kurtosis values. However, the values for skewness of all three indicators for Institutional Distance and several indicators for Industry Deviation are exceeding the critical values. However, since these variables are indicators for a latent variable and constructed in a similar pattern compared to the other indicators, transforming singular indicators is inadequate. In order to verify whether these variables will cause estimations issues, post-estimation tests need to be conducted.

In addition and related to all variables, in multivariate analyses, univariate indicators are not sufficient to assume the presence of normal distribution (Harlow, 2014). This is the case since *'univariate normality describes the distribution of only one variable in the sample while multivariate normality describes the joint distribution of all variables in the sample. The univariate normal distribution of each variable is a necessary, but not sufficient, condition for having a multivariate normal distribution'* (Gao et al., 2008, p. 3). Therefore, all variables need to be tested post-estimation for multivariate normality, whereas the previously mentioned variables require a closer examination.

Secondly, the residuals of the structural model need to be analysed. SEM assumes a consistent slope between each independent and dependent variable. As specified in the model, this relates to *GDP* and independent variables *Population, Wages, FDI* and *HumanCap*. Generating a regression including these variables enables graphically constructing the normality distribution of the residuals. These values are represented in (*Appendix D*). The graph indicates a normal distribution of the residuals of this regression. It can thus be stated that the observed variables collectively hold the assumption of linearity, eliminating biased estimates due to converging effects (Schimdt & Finan, 2018).

In addition, the observed regression and its error terms in the SEM need to be tested for heteroskedasticity. Heteroskedasticity refers to the random variability of standard errors, indicating a biased non-linear coefficient. Execution of the Breusch-Pagan test for heteroskedasticity shows that the null hypothesis of a constant variance needs to be rejected with a probability of 0.0309. In order to disregard biased estimates due to the lack of homogenous standard errors, the post-test variance-covariance matrix of the estimations needs to be inspected (StataCorp, 2013). Correlating variances of error terms need to be specified as covariances to ensure a sufficient model fit.

Thirdly, besides the variability of the error terms, (multi)collinearity between observed data needs to be considered. Multicollinearity generates an enhanced estimate between correlated relations, increasing the effect of a specific variable (Tarka, 2018). Whereas multicollinearity tests are only obtainable post-test since they are captured by the error terms, bivariate collinearities between exogenous, observed variables need to be tested ex-ante. Bivariate collinearities can be indicated by correlations in the SEM, and eliminate biased estimations on endogenous, unobserved *GDP*. Therefore, the Pearson correlation coefficients are estimated between the observed, exogenous variables, while the post-estimation variance-covariance matrix assesses multicollinearity between the latent variable indicators.

Table 4.8 represents Pearson's correlation coefficients between the exogenous, observed variables. To assign values to the magnitude of Pearson's correlation, Cohen's (1988) guidelines are adapted. He states that coefficient values '>.5' indicate strong correlation; '.3<.5' indicate medium correlation; and '.1<.3' indicate a small correlation. A strong correlation is present between *Human Capital, FDI* and *Population*. These correlations need to be implemented in the SEM.

Variable	LogHumanCap	LogFDI	Wages	Population
LogHumanCap	1			
LogFDI	0.811	1		
Wages	0.411	0.318	1	
Population	0.668	0.737	0.143	1

Table 4.8. *Pearson's correlation coefficients matrix of exogenous, observed variables.*

Variable Name	Description	Values	N	Mean	S.D.	Min	Max	Skewness	Kurtosis
CityID			209 (11)[1]						
Year		2000-2018	209 (19)[1]	2009	5.49	2000	2018		
GDP	Gross Domestic Product	CNY billion[2]	209	512.36	602.65	24.98	2524.99	1.653	4.773
LogGDP	*Logged GDP*		209	2.427	0.511	1.40	3.403	0.104	2.124
LogGDP_Lead	*Lead Logged GDP*		198	2.455	0.499	1.43	3.403	-0.331	2.105
Population	Total population	Total counts	209	5,408,171	3,373,686	427,782	14,904,400	0.575	2.814
Wages	Annual average wage	CNY[2]	209	53,880.18	38,069.96	8,545	178,387.5	1.151	3.778
HumanCap	New product output value	CNY 10.000[2]	209	17,035,384	25,334,320	67,328	128,420,224	1.993	3.803
LogHC	*Logged human capital*		209	6.619	0.880	4.828	8.109	-0.331	2.105
FDI	Foreign capital utilized	$10.000 US	209	319,519.7	561,051.6	2,584.97	5,145,641	5.12	37.085
LogFDI	*Logged FDI*		209	5.168	0.514	3.412	6.711	-0.415	4.078
Spatial connectivity	Bilateral connectivity to:	Minutes							
CHKSAR	*HKSAR*		190[3]	114	39.4	50.9	198.2	-0.385	2.544
CMSAR	*MSAR*		190[3]	83.7	42	4.6	162.6	-0.182	2.009
CGu	*Guangzhou*		190[3]	56.4	29.8	23.7	152.2	1.119	4.624
CSh	*Shenzhen*		190[3]	76.4	32.4	24.3	132.1	-0.285	1.945
CHu	*Huizhou*		190[3]	100.4	38.3	29.2	171.7	-0.464	2.238
CFo	*Foshan*		190[3]	107.8	41.2	28.6	198.2	-0.338	2.456
CZhu	*Zhuhai*		190[3]	85.4	41.5	4.6	162.6	-0.28	2.068
CDo	*Dongguan*		190[3]	69.9	26.5	23.7	112.3	-0.564	2.118
CZho	*Zhongshan*		190[3]	73.5	36.1	27.5	171.7	0.417	2.78
CJi	*Jiangmen*		190[3]	90	27.5	47.6	165.1	-0.358	3.447
CZha	*Zhaoqing*		190[3]	94.1	30.3	33.3	178.3	-0.444	3.331
Institutional distance	Index measure for:	Index 1-100							
LEGALID	*Legal distance*		209	60.90	10.21	51.58	84.43	1.61	3.71
NORID	*Normative distance*		209	68.02	5.86	63.46	83.13	1.54	3.70
CULTID	*Cultural distance*		209 (3)[1]	49.92	2.75	42.82	51.18	-1.834	4.637
Industry deviation	Bilateral deviation with:	% of GDP for:							
ManHKSAR	*HKSAR*	*Manufacturing*	190[3]	38.97	14.720	0.1	57.2	-1.366	4.075
ManMSAR	*MSAR*		190[3]	36.37	14.195	0.1	57.2	-1.224	3.724
ManGu	*Guangzhou*		190[3]	18.78	7.252	1.2	39.51	0.063	2.477
ManSh	*Shenzhen*		190[3]	15.02	13.492	0.1	50.7	1.116	2.901
ManZhu	*Zhuhai*		190[3]	14.44	15.323	0	47.7	1.128	2.732
ManFo	*Foshan*		190[3]	19.79	17.783	0	57.8	1.00	2.619
ManHu	*Huizhou*		190[3]	16.17	17.016	0.3	52	1.065	2.602

ManDo	*Dongguan*		190[3]	14.71	15.986	0	50.56	1.076	2.609
ManZho	*Zhongshan*		190[3]	17.20	17.392	0	56.7	1.036	2.550
ManJi	*Jiangmen*		190[3]	14.75	15.237	0	51.57	1.166	2.886
ManZha	*Zhoaqing*		190[3]	18.41	10.681	0	44.4	0.300	2.503
TerHKSAR	*HKSAR*	*Tertiary*	190[3]	43.26	15.936	0.1	61	-1.589	4.680
TerMSAR	*MSAR*		190[3]	43.26	15.927	0.1	61.8	-1.570	4.696
TerGu	*Guangzhou*		190[3]	21.34	7.830	3.44	39.36	0.072	2.207
TerSh	*Shenzhen*		190[3]	16.81	13.079	0.1	52.8	1.115	3.114
TerZhu	*Zhuhai*		190[3]	15.15	16.922	0	52.9	1.277	2.985
TerFo	*Foshan*		190[3]	18.02	19.652	0.1	60.2	1.116	2.701
TerHu	*Huizhou*		190[3]	19.54	19.839	0.2	61.8	1.137	2.764
TerDo	*Dongguan*		190[3]	15.51	15.552	0	53.23	1.227	3.016
TerZho	*Zhongshan*		190[3]	16.69	18.781	0	60.8	1.241	2.934
TerJi	*Jiangmen*		190[3]	16.23	18.583	0	58.46	1.202	2.844
TerZha	*Zhoaqing*		190[3]	16.88	18.796	0.1	59.8	1.200	2.860

Note. [1] *Total unique observations are in parenthesis if relevant;*
[2] *Determined using yearly average exchange rates;*
[3] *Observations for spatial connectivity and industry deviation with value 0 (non-bilateral observations) are excluded from descriptive statistics; included in analysis.*

Table 4.7. *Descriptions, descriptive statistics and values of the observed variables of interest for externality interdependencies*

5. Analysis and results

The following section focusses on the analysis of the constructed MSEM. Based on the methodology and operationalization, the analysed data is implemented in the model. Firstly, in order to derive results, the MSEM is specified and tested for general applicability. Since the data is formulated in such a way that only multigroup analysis generates valid results, these results are solely understood as eliminating identification issues of the model. Following is the multigroup analysis of the SEM, possible improvements to the data or model, and an extensive elaboration on all alternative estimation techniques applied to generate feasible estimations.

5.1. General model specifications

Firstly, in order to receive output from the SEM, the model needs to be identified. Whereas a normal equation needs enough identified observations of information, a structural equation model requires enough measured parameters to estimate the 'unknown' parameters (Harlow, 2014). Therefore, to provide a scale to the latent variables, at least one factor loading for each latent variable needs to be fixed. Which indicator is fixed does not alter any results. This results in six fixed factor loadings namely from *SC* to *CZha*; from *ID* to *LEGALID*; from *IndMan* to *ManZha*; from *IndTer* to *TerZha*; from *E37* to *ExterInter*; and from *E38* to *Lead_LogGDP.*

In addition to the identification of the factor loadings, the error terms of *ExterInter* and *Lead_LogGDP* are unidentified. This is the case as both variables are determined by unobserved variables and require at least one predetermined and observed variable. In order to resolve this issue, three possibilities are offered. The first and second possibility refer to determining the error variance (Oberski & Satorra, 2013). They state that the most recommended solution in SEM refers to fixing the error variance based on '*error variances of the indicators'* (p. 3), or based on error variances '*obtained from other sources such as published reliability studies'* (p. 4). Where the prior solution is not feasible in this structural equation model since both error variances are determined by multiple other latent variables (referring to a mixed subset of indicators), the latter solution is often considered. However, as elaborated previously, a comparable SEM study implementing path dependency and taking into account the dynamic interaction between the latent determinants of externality interdependencies is lacking. Therefore, error variances derived from other studies is no option. The third option refers to removing latent variable *ExterInter* from the SEM and directing each proposed causal effect directly on *Lead_LogGDP.* Since this latent variable was included in the SEM to rationalize the total effect of externality interdependencies, the total amount of relevant indirect and direct effects on *Lead_LogGDP* are now the unanimous effect of externality interdependencies on economic growth. Nevertheless, it is noteworthy that these effects are not without a nuance of interpretation since separate direct and indirect effects might be insignificant, while the total effect is significant on economic growth.

After identifying the model the previously estimated covariances between (modified) exogeneous variables are incorporated in the SEM framework. Subsequently, results can be estimated using ML estimations for fitting the saturated and independence models. The adapted path diagram, including standardized estimates without multigroup analysis, is represented in *figure 5.1*. With a total of 861 distinct sample observations and 944 parameters that need to be estimated, the general SEM framework has a degree of freedom of 767.

As elaborated in the methodology and operationalization section, the results of the SEM without multigroup analysis are conceptually unreliable as the data is specifically constructed for bilateral, between-groups analysis. This is also illustrated by the standardized estimated of *fig. 5.1*, which show conflicting between indicators and the relevant latent variables. For instance, *ManHKSAR* and *ManMSAR* appear to have a negative regression weight on latent variable *IndMan* since they have completely divergent industry structures compared to the other nine cities of the Guangdong province. Within the multigroup analysis, specific bilateral measures between HKSAR and the relevant cities exclusively capture these relations and thus referring to strong industry deviations and its relevant regressions weights. Therefore, the regression weights illustrated in *fig. 5.1* are not valid for interpretations or conclusions; multigroup analysis needs to be implemented. This also elaborates the high chi-square of 15.829 and CMIN/DF of 20,638, which illustrates the procedural invalidity of the SEM when not applying a multigroup analysis.

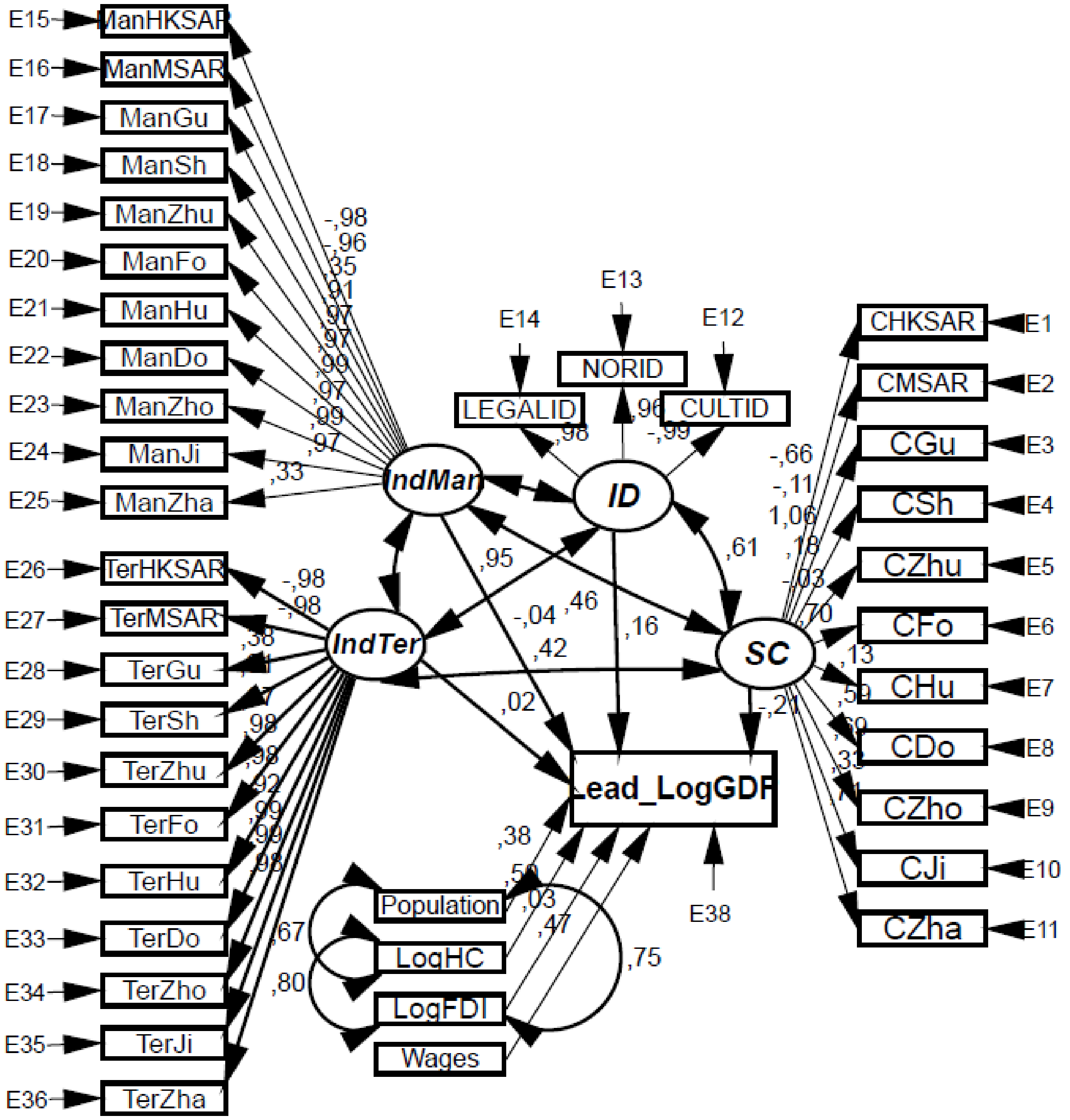

Figure 5.1. Path diagram of the estimated general structural equation model for externality interdependencies.

5.2. Multigroup analysis

The major target of the path diagram represented above is to assess variations across cities within the Greater Pearl River Delta of the parameters in this model by performing a multigroup analysis. As stated by Bou and Satorra (2010), '*multigroup models are valuable for an easily interpretable and detailed description of the variation at country level of the parameters characterizing the dynamic pattern'* (p. 745). Estimating a multigroup analysis and inspecting the estimations entails a multistep process.

First, several assumptions hold for MSEM. Amongst all cities, denoted by j, the path diagrams need to be similar. Since the data is constructed in such a matter that similar path diagrams are possible amongst the groups, this assumption holds. Secondly, the assessment of invariance across groups needs to be considered (Bou & Satorra, 2010; Ployhart & Oswald, 2004). Two model baselines for assessing invariances can be constructed. Referring to *eq. 2.6a*, it depends on whether all parameters are city-specific and assessed separately ($M0_{group}$) or are complete heterogeneous and based on multiple groups ($M1_{MG}$). Referring to the multiple groups' baseline model ($M1_{MG}$), '*various MG sub models are obtained by introducing constraints on these sets of parameters allowing us to assess country-level differences in the means and variances'* (Bou & Satorra, 2010, p. 745). However, since this study relies on city-specific data to measure the bilateral flow of externality interdependencies within the agglomeration, constraining parameters is conceptually invalid and generates biased results (as represented in *fig. 5.1)*. These constraints relate to equalling factor loadings and regression weights between latent and observed variables, as well as homogeneity between item intercepts of $M1_{MG}$. Therefore, several sub models of '*the hierarchical nested model approach frequently used to assess invariance across groups'* (Bou & Satorra, 2010, p. 751), cannot be considered in this study as the design of this study and data implies some constraints. Alternatively, an equivalent SEM multigroup analysis, or a configural model, will be performed in this study (Bou & Satorra, 2010). This configural model performs single-group analyses, calculating a factor structure for each group. Both considered multigroup models are represented in *table 5.1*.

$$\sum_{j=1}^{n} X_{ji} = \left(X_{1i,t}\,,\, X_{2i,t},\, X_{3i,t} \dots X_{ni,t} \right), \qquad j = 1,2,3,\dots n \qquad (2.6a).$$

Symbol	*Model Type*	*Model Constraints*	*Econometric Form (eq. 2.6a)*
$M0_{group1}$, $M0_{group1}$, $M0_{group\ etc.}$	Configural or equivalent forms models	Separate assessment of factor structure within each group	$M0_{Group1} = i\ (1)$; $M0_{Group1} = i\ (2)$; … $M0_{Group1} = i\ (n)$
$M1_{MG}$	Multiple groups baseline model	None save for identification; overall model fit across groups	$M1_{MG} = \sum i$

Table 5.1. *Description of multigroup baseline models considered. Adapted from Ployhart and Oswald (2004, p. 36).*

After identifying the model, the parameters of the structural equation model for all groups can be estimated using a configural forms model. Groups are denoted one till eleven, referring to respectively HKSAR, MSAR, Guangzhou, Shenzhen, Zhuhai, Foshan, Huizhou, Dongguan, Zhongshan, Jiangmen and Zhoaqing.

Regrettably, the structural equation model is unable to estimate parameters while performing a multigroup analysis in AMOS as it states that a non-positive definite sample matrix is present (see *appendix F)*. In order to derive multigroup estimates, a multistep analysis of possible estimation issues is conducted. Firstly, to eliminate any statistical problems caused by the software implemented in this study, the path diagram and estimates are tested using other SEM software. Estimating an identical path diagram for the MSEM in STATA results in no estimations. Therefore, unfortunately, it can be confirmed that the current dataset results in an improper solution for MSEM. There are several causes of improper solutions, specifically nonconvergence, empirical underidentification, or if the model is under identified (Kenny, 1979). As *fig. 5.1* indicates, it can be stated that the model is identified and converged since the ML process has generated estimations. Therefore, resolutions for empirical underidentification need to be considered.

The AMOS software highlights the issue of a non-positive definite sample matrix. This might be caused due to negative *p eigenvalues* for any of the matrices involved in ML estimations (Blunch, 2012). To elaborate issues and solutions regarding non-positive definite matrices, Blunch (2012) formulates the following formula:

$$F = f(S - \Sigma(\hat{\partial})) \qquad (5.1),$$

where *F* represents an econometric function for structural equation model estimations, consisting of the sample covariance matrix (S) and the largest probability (in ML) of reproducing these covariances using the estimated parameters ($\hat{\partial}$). If AMOS indicates that there is a non-positive sample matrix, it is definite that *S,* the sample covariance matrix, requires a closer examination. If the sample covariance matrix is non-positive definite, it is often caused by data entry errors or pairwise deletion, disturbing linearity or correlation coefficients between observed variables, or issues regarding sample size (Blunch, 2012). As *table 4.7* indicates, as well as the pre-test data analysis, data is consistent and non-missing. In addition, previous outlier analysis and appropriate log-transformations have eliminated random measurement errors and extreme variances. Therefore, it is assumed that empirical underidentification might be an issue of linearity amongst observed variables, strong correlation coefficients or the total sample size. These issues will be discussed sequentially while challenging the MSEM after each revision.

Primarily, linearity amongst the observed variables needs to be discussed. Blunch (2012) states that:

> *'due to rounding errors, strong co-linearity among manifest variables will lead to the same problems as if this matrix was semi-definite, i.e. it will result in empirical underidentification…. So be prepared to face problems if you have correlations larger than 0.80 among the manifest variables'.* (p. 31).

As *table 4.8* illustrates, *LogFDI* and *LogHumanCap* have a correlation coefficient exceeding Blunch's critical value of 0.8. Therefore, it is reasonable to exclude *LogHumanCap* from the control variables. Considering the observed variables for each indicator, *Appendix E* (*table E1, E2,* and *E3*) shows the correlation coefficients between the observer variables for each latent factor. The coefficients between the indicators for spatial connectivity are not exceeding the critical value of 0.8, while the values for industry deviation are highly correlated and often reaching values between 0.95 and 0.8.

Whereas these high correlation coefficients are somehow expected due to the operationalization of the variable for industry similarity, several revisions need to be considered in order to provide valid MSEM estimations. Since each variable indicates bilateral industry similarity, it is logical that comparable cities have greater correlations with each other compared to diverging cities. Therefore, *tables E2* and *E3* should not be considered as main measures for correlation. Taking into account the correlation coefficients between the measures for each group is more indicating for empirical underidentification in MSEM. This results in eleven correlation coefficient tables which include only correlations regarding a specific group. These correlation tables for manufacturing and tertiary industry deviation show less disturbing results, namely respectively a total of four and ten coefficients exceeding the critical term of 0.8 for HKSAR, two and seven coefficients for MSAR, three and five for Guangzhou, seven and thirteen for Shenzhen, two and zero for Zhuhai, sixteen and eleven for Foshan, seven and three for Huizhou, six and five for Dongguan, two and zero for Zhongshan, two and nine for Jiangmen, and 32 values for manufacturing and 12 for tertiary that exceed the critical point of 0.8 for Zhaoqing. Obviously these grouped correlation coefficients are still present in abundance and varied amongst all groups. Whereas eliminating the specific variables that have linearity is the most obvious revision, it is not an option in this study as each variable represents a specific bilateral relation. Therefore, in order to generate multigroup estimations of the SEM, industry deviation is left out of the analysis. Unfortunately, no multigroup estimations can be generated after eliminating *LogHumanCap* and the measurement model for industry similarity.

In addition, in MSEMs, several parameters can be fixed between equivalent models. As elaborated before due to the bilateral perspective of externality interdependencies, fixing parameters between latent measurement structures and the structural model generate conceptually invalid results. However, it is theoretically just to assume generalizable factors between the specified control variables (*Population, LogHC, LogFDI,* and *Wages*) and dependent variable *Lead_LogGDP*. As elaborated in the methodology, all relationships amongst these variables are presumed in agglomeration studies as well as the correlations presented in *table 4.8.* Unfortunately, after anchoring the parameters related to the measurement construct of the cited variables for each group, no MSEM estimations can be generated as the same error message in AMOS appears (see *Appendix F*).

Secondly, the total sample size in this study needs to be considered. As stated by Newsom (2020), in the case of '*empirical underidentification, the model overall has positive degrees of freedom, but there is insufficient covariance information in a portion of the model for the computer to generate valid estimates'* (p. 1). The sample size is often a widely debated subject, with proposed minimal sample sizes differing enormously between studies (Blunch, 2012; Hox & Bechger, 1998; Hox & Maas, 2001; Kline, 2010). Nevertheless, the effect of small sample sizes on parameters of a SEM is widely agreed upon. Without the presence of an adequate amount of covariance information, estimations for parameters, chi-square tests, overall fit measures and error covariances become unmeasurable and invalid (Kyriazos, 2018). This is also stated by Blunch (2012), who states that '*it should be pointed out that any of the three matrices* [of which two are represented in *eq. 5.1*] *could cause problems if the sample is too small'* (p. 101). Thus, even if linearity issues, error messages and the validity of the estimations methods are considered, data estimations can be severely influenced by a lack of sample size. Whereas a discussion related to the applied sample size in this study and the contradiction between generalization and specification is broadly present in the discussion of this paper, the multigroup sample sizes need to be reconsidered to attempt to generate valid estimates.

In order to improve the sample size, several group structures are principally established and implemented in the MSEM. *Table 5.2* represents the considered groups as well as the theoretical reasoning behind the group formation and the number of observations. It is noteworthy that these groups do not hold a theoretical motivation; they are virtuously created in order to generate estimates in the MSEM by gradually increasing sample size. All groups represented in *table 5.2* are gradually tested in the adapted multigroup structural equation model. Regrettably, no valid estimates can be generated due to similar issues of empirical underidentification.

Group constraint	***Group A***	***Group B***	***Group C***
Geography[1]	MSAR, Zhongshan, Jiangmen, Zhaoqing, Zhuhai *(90)*	Guangzhou, Dongguan, Foshan *(54)*	HKSAR, Huizhou, Shenzhen *(54)*
Colonial past	HKSAR, MSAR (36)	Guangzhou, Shenzhen, Dongguan, Foshan, Huizhou, Zhuhai, Zhaoqing, Zhongshan, Jiangmen *(162)*	
Core-periphery[2]	HKSAR, Shenzhen, Guangzhou *(54)*	Dongguan, Foshan, Huizhou, Zhuhai, Zhaoqing, Zhongshan, Jiangmen, MSAR *(144)*	

Note. [1] *Partially based on Hou and Li's (2011) spatial division of the Greater Pearl River Delta.*
[2] *Based on economic spatial forces calculated by Yang et al. (2018).*

Table 5.2. *Alternative MSEM group constraints with corresponding sample sizes in parenthesis.*

In addition, alternative estimation methods need to be discussed. Alternative estimation methods are often considered in small sample data analyses and in the case of empirical underidentification (Kline, 2010; McNeish, 2016).

First, Bayesian estimation methods are scrutinized. McNeish (2016) states that '*although Bayesian methods are better equipped to model data with small sample sizes, estimates are highly sensitive to the specification of the prior distribution'* (p. 750). These prior values provide additional information variances besides the regular database, providing a more stable base for small size SEM analysis. Whereas several studies have used statistical measures and parameters from previous studies to calculate informative or subjective priors in small sample size (multigroup) SEMs, estimates about prior parameters values for a comparable study are lacking. Several researchers created uninformative priors (or Jeffrey's Priors) based on their theoretical framework. However, Bayesian estimates are exceedingly dependent on these priors and improperly estimated priors can generate strongly biased estimates. This issue with informative priors is also elaborated by McNeish (2016), who states that:

> *With small sample data, the prior distribution is given more weight in calculating the posterior distribution, relative to whether the data came from a larger sample. … Although this seems trivial, it can have enormous implications with small samples, as the posterior distribution is highly reliant on how the prior was specified.* (p. 752)

As the experimental framework and bilateral perspective of this study are progressive and eliminate the possibility of using pre-existing values to determine prior distribution, an extensive theoretical and empirical analysis to determine prior parameters is required. Therefore, using Bayesian estimations is considered as a feasible alternative to overcome empirical underidentification with ML-estimations. However, an extensive theoretical analysis is required to avoid strongly biased estimates.

Alternatively, several other estimations methods in MSEM are considered. Where weighted least squares (WLS) and generalized least squares (GLS) only generate suitable and valid estimates for relatively large sample sizes, unweighted least squares (ULS) estimations are reliable with a small sample size (Schermelleh-Engel et al., 2003). However, ULS estimations need to be interpreted with vigilance as estimates are easily biased or overemphasized. Unfortunately, no valid estimates can be generated using ULS and the prespecified MSEM alterations.

Consequently, since the provisional and general structural equation model represented in *figure 5.1* has generated (conceptually invalid) estimations, it can be assumed that the path diagram is conceptually valid. However, the MSEM needs to be rejected due to empirical underidentification of the selected data. The following section broadly analyses and discusses the operationalization and methodology applied in this study, as well as recommendations for future implementations and applications.

6. Conclusion and discussion

Building upon the notion of externality interdependencies, this study has specified the relationship between externalities and path dependency in agglomerations. The supportive relationship between Marshall- and Jacobs externalities has been qualified, contributing to the general understanding in economic geography that agglomerations yield positive externalities. Though, externalities are highly dependent on place specific developments. This path dependency of externalities, referring to endogenous, dynamic and historical developments within the agglomeration, results in highly embedded and integrated agglomeration externalities.

The resourceful concept of externality interdependencies allows to theoretically, econometrically and empirically analyse the path dependent role of bilateral externalities for the Greater Pearl River Delta from 2000 till 2018. Firstly, based on prior theoretical studies the dynamic role of several dimensions and relevant concepts in agglomeration studies are analysed. The theoretical framework provides strong assumptions that spatial connectivity and institutional distance negative effect the presence of externalities within agglomerations, while industry similarity has a positive effect. In addition, it is assumed that the prior two have a supportive role on externalities in the case of a local integrated economy.

At a more practical level, the econometric representations of externality interdependencies enable one to empirically test the implications of this latest concept. Using several (local) databases to operationalize the specified variables for the GPRD from a bilateral perspective, an experimental multigroup structural equation model is suggested in this study. The results show that, given the multigroup SEM approach, no estimations are feasible. Since invalid estimations are retrievable when considering the total sample size, and thus eliminating the bilateral MSEM structure, empirical underidentification is the main reason for improper solutions. Several resolutions are proposed to eliminate the non-positive definite matrices, including data adaptations, alternative estimations methods and various group analyses to increase relative sample sizes. Unfortunately, since no valid MSEM estimations can be derived, a more extensive discussion on the operationalization and methodology strategy is necessary.

Specifically, in the case of empirical underidentification, the problem is a result of the data and not of the model (Blunch, 2012). Issues regarding the data are often difficult to find, and after the resolutions proposed in the previous section, the data collection needs to be analysed. Since this study implements an experimental approach, no comparable studies using similar variables are present. Therefore, the most suited bilateral measures for institutional distance, spatial connectivity and industry similarity are considered in this study. Where on the one side these measures have the highest validity and reliability, within a bilateral MSEM framework they might have contributed to the empirical underidentification. Explicitly, the three implemented latent variables have indicators with a strong sense of serial correlation and linearity. Several pre-estimate tests have shown that these issues should not be an issue as several critical values are not being reached. In addition, as shown in *fig. 4.7*, variables are considered to be normally distributed. Nevertheless, empirical underidentification is often a combined result of several (minor) data challenges which cannot always be sighted for pre-test or post-test (Blunch, 2012). Therefore, serial correlation, linearity, or other statistical issues, should be considered as factors that have resulted in empirical underidentification. In order to overcome these concerns, alternative operationalization options should be considered. Operationalizing alternative indicators for the three latent structures in this study require a broad (theoretical) analysis as the most academically verified indicators are applied in this study. An alternative approach would be to first consider each latent variable in a bilateral measure individually. Therefore, it is suggested when considering a similar MSEM structure as implemented in this study, an extensive and gradual theoretical and empirical approach should be designed.

In addition, the data itself should be discussed. As stated in the descriptive statistics and data analysis section, *GDP* and *FDI* have some outliers that are not eliminated due to the statistical power of these outliers. Besides these outliers, numerous remarkable data entries can be found in the statistical yearbooks. Values for economic growth (*GDP*) and foreign direct investments (*FDI*) often reach vast values. For instance, the Special Administrative Region of Macau experienced twice a GDP growth of over 30%, while Foshan, Zhaoqing and Zhongshan experienced values greater than 25%. Several extremes in the annual increase of *FDI* are also present with a growth of 595% as the supreme value. As these values are collected from national and regional statistical yearbooks and shouldn't be up for any considerations, specific sample sizes might experience a certain degree of heteroskedasticity.

Lastly, the fundamental sample size in this study needs to be debated. An inappropriate sample size is directly related to empirical underidentification (Kyriazos, 2018; Newsom, 2020). This study involved a total sample size of 209 which can be considered as limited for a MSEM. The selection of observations is determined by the case study, availability of significant (panel) data, and the group considerations. The Greater Pearl River Delta has provided an appealing perspective on the dynamic relations of the specified variables. However, due to the selection of this specific region, the total sample size was limited to eleven prefecture-level cities and a time series of eighteen years due to inconsistent measurements. The bilateral MSEM framework diminished group samples drastically. However, since the alternative group structures have not generated valid estimates, the total sample size is up for considerations.

These considerations refer to the issue between specialization and generalization. Where in economic studies generalization is a desired empirical outcome, the inclusion of path dependency in externality interdependencies eliminates this objective. Path dependency requires a specific case analysis to be studied, involving all historical and dynamic developments to verify the degree of externalities. In contrast to agglomeration studies aiming for generalizable results, estimations for externality interdependencies can only be generalized after careful considerations. This contradiction between specialization and generalization is directly related to the sample size for externality interdependencies. Whereas generalization and externalities, in general, allow a broad inclusion of cases and observations, a specified region and externality interdependencies require a specific case study which is limited by the availability of data. On the one side, the specialization of externality interdependencies allows measuring specific statistics and developments, while, on the other side, these statistics only comprehend the specific subject of study. Discussions regarding this issue have been prevalent for years, to which the notion of externality interdependencies is a new-found contribution.

Besides the operationalisation of the variables, the applied methodology needs to be discussed. A multigroup structural equation model is applied in order to specify bilateral relations instead of generalizing dynamic relations or sums of externality interdependencies. Whereas the aim of this study, represented in the constructed hypotheses, is to measure city-specific externality interdependencies for regional policy- and localization strategies, a MSEM is the appropriate statistical procedure. Nevertheless, if future studies on externality interdependencies aim to measure more generalizable, variable dependent, unilateral or multilateral externalities interdependencies, other statistical procedures can be more satisfying. For instance, multilevel ML or factor score analysis estimations. The possibility of alternative estimation methods are both an opportunity and a vulnerability of the developed framework in this study. Where, on the one hand, the theoretical and econometric framework for externality interdependencies allows an abundance of future implications contingent on the research scope and objective. On the other hand, a strong recommendation for a methodology and empirical procedure for measuring externality interdependencies is lacking according to this study. Nevertheless, it is believed that the econometric framework provides a substantial base in order to develop the most practicable and appropriate statistical model to qualify externality interdependencies within agglomerations.

So, regarding the proposed hypotheses, it is self-evident that they cannot be empirically accepted or rejected as valid estimations are imperative. Nevertheless, the theoretical framework and assumptions advance towards a specific direction which requires a broad analytical and statistical procedure. This also refers to the research questionof whether externality interdependencies can be measured by a MSEM. Where the absence of statistical outcomes indicate that it is not feasible, this assumption needs to be considered with caution. As the applied operationalization and methodology in this study for the MSEM do not generate estimations, alternative statistical procedures for a MSEM might generate feasible results. For instance, when operationalizing the data in a singular and comparable setting, compared to a bilateral approach, might eliminate any of the previously elaborated issues. In this case, a MSEM can be the appropriate statistical procedure to measure externality interdependencies.

To this extent and despite the absence of valid MSEM estimations, the implications of the framework for externality interdependencies for either agglomeration- and (economic) geography studies, as well as regional policies are clear. First, regional economies can be visualized and understood from a more integrated perspective, while taking into account the extensive role of local developments. The specifications of externality interdependencies are an innovative tool to visualize the effects of specific policies on existing processes and methods. Second, the theoretical framework provides an extensive base for future implications and locational studies. This base allows viewing local agglomerations from a new perspective in which agglomeration benefits can be investigated within a dynamic framework. This framework can be adapted according to the variables of interest, as well as the underlying connections between these dimensions.

Nevertheless, for future implementations, the path dependent perspective requires an extensive historical, institutional and developmental analysis to correctly visualize and understand the underlying structure of an agglomeration. Referring to the contradiction between specialization and generalization, future implications are unlimited when considering agglomerations as autonomous entities. The main objective for future implementations of externality interdependencies thus remains invariable: obtaining the equilibrium between a specialization and generalization strategy for externality interdependencies.

7. References

Alexander, E. A. (2012). The effects of legal, normative, and cultural-cognitive institutions on innovation in technology alliances. *Management International Review, 52*(6), 791-815.

Allison, P. D. (2003). Missing data techniques for structural equation modeling. *Journal of abnormal psychology, 112*(4), 545.

Amin, A. (1999). An institutionalist perspective on regional economic development. *International journal of urban and regional research, 23*(2), 365-378.

Amiti, M., & Cameron, L. (2007). Economic geography and wages. *The Review of Economics and Statistics, 89*(1), 15-29.

Arrow, K. J. (1962). The economic implications of learning by doing. In *Readings in the Theory of Growth* (pp. 131-149). Palgrave Macmillan, London.

Artal-Tur, A., Navarro-Azorín, J. M., García-Sánchez, A., & Alamá-Sabater, M. L. (2012). The role of destination spatial spillovers and technological intensity in the location of manufacturing and services firms. *Environment and Planning B: Planning and Design, 39*(6), 991-1005.

Audretsch, D. B., & Feldman, M. P. (2004). Knowledge spillovers and the geography of innovation. In *Handbook of regional and urban economics* (Vol. 4, pp. 2713-2739). Elsevier.

Bae, J. H., & Salomon, R. (2010). Institutional distance in international business research. *Advances in international management,* 23, 327-349.

Bay Area. (n.d.). Retrieved June 23, 2020, from https://www.bayarea.gov.hk/en/home/index.html

Bie, J., de Jong, M., & Derudder, B. (2015). Greater Pearl River Delta: historical evolution towards a global city-region. *Journal of Urban Technology, 22(*2), 103-123.

Blunch, N. (2012). *Introduction to structural equation modeling using IBM SPSS statistics and AMOS*. Sage.

Bou, J. C., & Satorra, A. (2010). A multigroup structural equation approach: A demonstration by testing variation of firm profitability across EU samples. *Organizational Research Methods, 13*(4), 738-766.

Braunerhjelm, P. (2011). Entrepreneurship, innovation and economic growth: interdependencies, irregularities and regularities. *Handbook of Research on Innovation and Entrepreneurship*. Elgar: Cheltenham, 161-213.

Burger, M. J., & Meijers, E. J. (2016). Agglomerations and the rise of urban network externalities. *Papers in Regional Science, 95*(1), 5-15.

Burnette, J. L., & Williams, L. J. (2005). Structural equation modeling (SEM): An introduction to basic techniques and advanced issues. *Research in organizations: Foundations and methods of inquiry*, 143-160.

Byrne, B. M. (2010). Structural equation modelling with AMOS: basic concepts, applications, and programming (2nd ed.). *New York, NY*: Routledge.

Camagni, R., Capello, R., & Caragliu, A. (2017). Static vs. dynamic agglomeration economies: Spatial context and structural evolution behind urban growth. In *Seminal studies in regional and urban economics* (pp. 227-259). Springer, Cham.

Capello, R. (2000). The city network paradigm: measuring urban network externalities. *Urban Studies, 37*(11), 1925-1945.

CBRE Research (2016). Greater Pearl River Delta Infrastructure Outlook: Real Estate Development Driven by Reducing Distance between Cities. *CBRE Greater China Research Team.*

Census and Statistics Department. (2019). Hong Kong Annual Digest of Statistics (various years) [data set]. *The Government of the Hong Kong Special Administrative Region, Hong Kong, China.* Retrieved July 29, 2020, from https://www.censtatd.gov.hk.

Chen, P. (2019, October 25). Underused mega-bridge hungry for traffic. *Asia Times.* https://asiatimes.com/2019/10/underused-mega-bridge-hungry-for-traffic-as-competition-looms/

Cho, K. R., & Padmanabhan, P. (2005). Revisiting the role of cultural distance in MNC's foreign ownership mode choice: the moderating effect of experience attributes. *International Business Review, 14*(3), 307-324.

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences.* Academic press.

Dauth, W. (2010). Agglomeration and regional employment growth (No. 7/2010). *IAB-Discussion Paper.*

David, P.A. (1985). Clio and the Economics of QWERTY. *American Economic Review*, 75, p. 332-337

De Groot, H. L., Poot, J., & Smit, M. J. (2009). Agglomeration externalities, innovation and regional growth: theoretical perspectives and meta-analysis. *Handbook of regional growth and development theories*, 256-281.

De Lucio, J. J., Herce, J. A., & Goicolea, A. (2002). The effects of externalities on productivity growth in Spanish industry. *Regional Science and Urban Economics, 32*(2), 241-258.

Delgado, M., Porter, M. E., & Stern, S. (2016). Defining clusters of related industries. *Journal of Economic Geography, 16(*1), 1-38.

Deloitte. (2013). Research Report on Investment Environment of Jiangmen City, Guangdong Province, China. *Deloitte Touche Tohmatsu Limited*, 1-50.

Desjardins, E. (2011). Historicity and Experimental Evolution. *Biology and Philosophy 26*, 339–64.

Desrochers, P., & Sautet, F. (2008). "Entrepreneurial Policy: The Case of Regional Specialization vs. Spontaneous Industrial Diversity," *Entrepreneurship: Theory & Practice*, 32, 813–832.

Di Clemente, R., Chiarotti, G. L., Cristelli, M., Tacchella, A., & Pietronero, L. (2014). Diversification versus specialization in complex ecosystems. *PloS one, 9(*11).

Dicken, P. (2015). *Global shift: Mapping the contours of the changing world economy.* SAGE Publications Limited.

Djelic, M. L., & Quack, S. (2007). Overcoming path dependency: path generation in open systems. *Theory and society, 36*(2), 161-186.

Dogaru, T., Van Oort, F., & Thissen, M. (2011). Agglomeration economies in European regions: perspectives for objective 1 regions. *Tijdschrift voor economische en sociale geografie, 102*(4), 486-494.

Ejermo, O. (2005). Technological diversity and Jacobs' externality hypothesis revisited. *Growth and Change, 36(*2), 167-195.

Ereshefsky, M. (2014). Species, historicity, and path dependency. *Philosophy of Science, 81*(5), 714-726.

Feld, W. J., & Boyd, G. (Eds.). (2016). *Comparative regional systems: West and East Europe, North America, the Middle East, and Developing Countries.* Elsevier.

Fu, S., & Hong, J. (2011). Testing urbanization economies in manufacturing industries: urban diversity or urban size?. *Journal of Regional Science, 51*(3), 585-603.

Gao, S., Mokhtarian, P. L., & Johnston, R. A. (2008). Nonnormality of data in structural equation models. *Transportation Research Record, 2082*(1), 116-124.

Gaur, A. S., & Lu, J. W. (2007). Ownership strategies and survival of foreign subsidiaries: Impacts of institutional distance and experience. *Journal of management, 33*(1), 84-110.

Gertler, M. S. (2005). Tacit knowledge, path dependency and local trajectories of growth. In *Rethinking Regional Innovation and Change* (pp. 23-41). Springer, New York, NY.

Gerritse, M., & Arribas-Bel, D. (2018). Concrete agglomeration benefits: do roads improve urban connections or just attract more people?. *Regional Studies, 52*(8), 1134-1149.

Giuliani, E. (2003). Knowledge in the Air and its Uneven Distribution: A story of a Chilean Wine Cluster. Presented at the *DRUID Winter Conference 2003,* Aalborg: Denmark

Glaeser, E. L., Kallal, H. D., Scheinkman, J. A., & Shleifer, A. (1992). Growth in cities. *Journal of political economy, 100*(6), 1126-1152.

Guangdong Statistical Bureau. (2019). Statistical Yearbook of Guangdong (various years) [data file]. *China Statistics Press,* Beijing, China. Retrieved July 28, 2020, from stats.gd.gov.cn/gdtjnj/

Han, Y., & Jia, H. (2017). Simulating the spatial dynamics of urban growth with an integrated modeling approach: A case study of Foshan, China. *Ecological Modelling,* 353, 107-116.

Hao, Z. (2011). Macau history and society. *Hong Kong University Press.*

Harlow, L. L. (2014). *The essence of multivariate thinking: Basic themes and methods*. Routledge.

Henderson, V. (1997). Externalities and industrial development. *Journal of urban economics, 42*(3), 449-470.

HKTDC Research. (2015a). Huizhou Dayawan Economic and Technological Development Zone. *Retrieved from https://hkmb.hktdc.com/en/1X09WO7T/hktdc-research/Huizhou-Dayawan-Economic-and-Technological-Development-Zone-including-Huizhou-Export-Processing-Zone*

HKTDC Research. (2015b). Zhongshan (Guangdong) City Information. *Retrieved from https://hkmb.hktdc.com/en/1X09VT9C/hktdc-research/Zhongshan-Guangdong-City-Information*

Hofstede, G. (2001). Culture's consequences: Comparing values, behaviors, institutions, and organizations across nations. *Thousand Oaks*, CA: Sage Publications.

Hong, S., & Xiao, Y. (2016). The influence of multiple specializations on economic performance in US Metropolitan Areas. *Sustainability, 8*(9), 963.

Hongbin, Q., & Lam, K. (2018). China's Greater Bay Area: Many Cities, One Goal. *HSBC Global Research.*

Hotho, J. J. (2009). A measure of comparative institutional distance. *Center for Strategic Management and Globalization*, 14.

Hou, Q., & Li, S. M. (2011). Transport infrastructure development and changing spatial accessibility in the Greater Pearl River Delta, China, 1990–2020. *Journal of Transport Geography, 19*(6), 1350-1360.

Howitt, P., & Aghion, P. (1998). Capital accumulation and innovation as complementary factors in long-run growth. *Journal of Economic Growth, 3*(2), 111-130.

Hox, J. J., & Bechger, T. M. (1998). An introduction to Structural Equation Modeling. *Family Science Review,* 11, 354-373.

Hox, J.J. & Maas, C.J.M. (2001). The Accuracy of Multilevel Structural Equation Modeling With Pseudobalanced Groups and Small Samples. *Structural Equation Modeling*, 8, 2, 157-174

Huesmann, L. R., & Guerra, N. G. (1997). Children's normative beliefs about aggression and aggressive behavior. *Journal of personality and social psychology, 72*(2), 408.

Hui, E. C., Li, X., Chen, T., & Lang, W. (2018). Deciphering the spatial structure of China's megacity region: A new bay area—The Guangdong-Hong Kong-Macao Greater Bay Area in the making. *Cities.*

Huifeng, H. (2018, January 12). *Will 2018 be the year Shenzhen and Guangzhou finally overtake Hong Kong? South China Morning Post.* https://www.scmp.com/news/china/economy/article-/2127862/will-2018-be-year-shenzhen-and-guangzhou-finally-overtake-hong

Illy, A., Hornych, C., Schwartz, M., & Rosenfeld, M. T. (2009). Urban growth in Germany–The impact of localization and urbanization economies (No. 19/2009). *IWH Discussion Papers.*

Ionascu, D., Meyer, K. E., & Estrin, S. (2004). Institutional distance and international business strategies in emerging economies. *William Davidson Institute Working Paper*, 728, 2-41.

Jackson, R. (2015). Are industry clusters and diversity strange bedfellows?.

Jacobs' Externalities: Where We Have Been and Where We Might Go in Studying How Urbanization Externalities Affect Innovation (2017). *Coins17*, Detroit. Retrieved from http://detroit17.coinsconference.org/papers

Jones, C. I. (2019). 'Paul Romer: Ideas, nonrivalry, and endogenous growth'. *Scandinavian Journal of Economics,* 121, 859–83.

Jones, G., & Lundan, S. (2001). The ''commonwealth effect'' and the process of internationalisation. *The World Economy, 24*(1), 99–118.

Kenny, D. A. (1979). *Correlation and causality*. New York: Wiley

Khanna, P., & Sevcik, T. (2012, 21 June). 'One Mega-City, Many Systems': The Evolution of Hong Kong. *Bloomberg*. Retrieved from https://www.bloomberg.com/news/articles/2012-06-21/-one-mega-city-many-systems-the-evolution-of-hong-kong

Klenow, P. J., & Rodriguez-Clare, A. (2005). Externalities and growth. *Handbook of economic growth,* 1, 817-861.

Kline, R. B. (2010). Promise and pitfalls of structural equation modeling in gifted research. *American Psychological Association*, p. 147–169.

Krugman, P. (1991). Increasing returns and economic geography. *Journal of political economy, 99*(3), 483-499.

Krugman, P. R., & Obstfeld, M. (2006). International economics: Theory and Policy. *Boston, MA: Addison Wesley.*

Kumar, N., & Pradhan, J. P. (2005). Foreign direct investment, externalities and economic growth in developing countries: Some empirical explorations. In *Multinationals and foreign investment in economic development* (pp. 42-84). Palgrave Macmillan, London.

Kumssa, A., & Mbeche, I. M. (2004). The role of institutions in the development process of African countries. *International Journal of Social Economics.*

Kyriazos, T. A. (2018). Applied psychometrics: sample size and sample power considerations in factor analysis (EFA, CFA) and SEM in general. *Psychology, 9*(08), 2207.

Lee, S., Guo, W. J., Tsang, A., Mak, A. D., Wu, J., Ng, K. L., & Kwok, K. (2010). Evidence for the 2008 economic crisis exacerbating depression in Hong Kong. *Journal of affective disorders, 126(*1-2), 125-133.

Li, X., Hui, E. C. M., Lang, W., Zheng, S., & Qin, X. (2020). Transition from factor-driven to innovation-driven urbanization in China: A study of manufacturing industry automation in Dongguan City. *China Economic Review,* 59, 101382.

Li, K., Li, Q., Zhou, X. G., & Fan, Z. (2015). Durability design of the Hong Kong–Zhuhai–Macau sea-link project: Principle and Procedure. *Journal of Bridge Engineering, 20*(11), 04015001.

Liebowitz, S. J., & Margolis, S. E. (1995). Path dependence, lock-in, and history. *Journal of Law, Economics, & Organization*, 205-226.

Liu, Y. (2019, 18 February). Can China's Greater Bay Area really rival the likes of Tokyo, New York and San Francisco*? South China Morning Post.* https://www.scmp.com/business/china-business/article/2186681/can-chinas-greater-bay-area-really-rival-likes-tokyo-new

Lu, W., & Tam, V. W. (2013). Construction waste management policies and their effectiveness in Hong Kong: A longitudinal review. *Renewable and sustainable energy reviews, 23,* 214-223.

Malizia, E. & Feser, E. (1999) Understanding Local Economic Development*. CUPR* Press/Transaction Press: Piscataway, NJ.

Marll, S. D. (2008). industry structure similarities, trade Agreements, and business cycle synchronization. *Gettysburg Economic Review, 2*(1), 4.

Marshall, A. (2009). Principles of economics: unabridged eighth edition. *Cosimo, Inc.* (Original work published in 1920)

Martin, R. (2011). 15 Regional economies as path-dependent systems: some issues and implications. *Handbook of regional innovation and growth,* 198.

Martin, R., & Sunley, P. (2006). Path dependence and regional economic evolution. *Journal of economic geography, 6*(4), 395-437.

McCann, P., & Acs, Z. J. (2011). Globalization: countries, cities and multinationals. *Regional Studies, 45(*1), 17-32.

McCann, P., & Van Oort, F. (2019). Theories of agglomeration and regional economic growth: a historical review. *Handbook of regional growth and development theories*. Edward Elgar Publishing.

McNeil, C. R., & Moore, W. T. (2005). Dismantling internal capital markets via spinoff: effects on capital allocation efficiency and firm valuation. *Journal of Corporate Finance, 11*(1-2), 253-275.

McNeill, D. (2016). Governing a city of unicorns: technology capital and the urban politics of San Francisco. *Urban geography, 37*(4), 494-513.

McNeish, D. (2016). On using Bayesian methods to address small sample problems. *Structural Equation Modeling: A Multidisciplinary Journal, 23*(5), 750-773.

Meyer, S., & Revilla Diez, J. (2015). One country, two systems: How regional institutions shape governance modes in the greater Pearl River Delta, China. *Papers in Regional Science, 94*(4), 891-900.

Neffke, F., Henning, M., Boschma, R., Lundquist, K. J., & Olander, L. O. (2011). The dynamics of agglomeration externalities along the life cycle of industries. *Regional studies, 45*(1), 49-65.

Newsom, J. T. (2020). Improper Solutions in SEM [Online Handout in Psy 523/623]. Department of Psychology, Portland State University, Portland, Oregon.

Ng, M. K. (2003). Shenzhen. *Cities, 20*(6), 429-441.

North, D. C. (1991). Institutions. *Journal of economic perspectives, 5*(1), 97-112.

Oberski, D. L., & Satorra, A. (2013). Measurement error models with uncertainty about the error variance. *Structural Equation Modeling: A Multidisciplinary Journal, 20*(3), 409-428.

Outline Development Plan for the Guangdong-Hong Kong-Macao Greater Bay Area. (2019, February 18). Retrieved from https://www.bayarea.gov.hk/filemanager/en/share/pdf/Outline_ Development_ Plan.pdf

Pessoa, A. (2011). Externalities, clusters and economic growth: The cluster policy paradox.

Phelps, N. A. (1992). External economies, agglomeration and flexible accumulation. *Transactions of the Institute of British Geographers*, 35-46.

Phelps, N. A., Fallon, R. J., Williams, C. L. (2001). Small firms, borrowed size and the urban-rural shift. *Regional Studies* 35: 613-624

Pike, A., Marlow, D., McCarthy, A., O'Brien, P., & Tomaney, J. (2015). Local institutions and local economic development: the Local Enterprise Partnerships in England, 2010–. *Cambridge Journal of Regions, Economy and Society, 8*(2), 185-204.

Ployhart, R. E., & Oswald, F. L. (2004). Applications of mean and covariance structure analysis: Integrating correlational and experimental approaches. *Organizational Research Methods, 7*(1), 27-65.

Porter, M. E. (1990). The competitive advantage of nations. *Harvard business review, 68*(2), 73-93.

Porter, M. E. (1998). Clusters and the new economics of competition (Vol. 76, No. 6, pp. 77-90). Boston: *Harvard Business Review.*

Preen, M. (2018a). China's Mega City Clusters: Jing-Jin-Ji, Yangtze River Delta, Pearl River Delta. *China Briefing*. https://www.china-briefing.com/news/chinas-mega-city-clusters-jing-jin-ji-yangzte-river-delta-pearl-river-delta/

Preen, M. (2018b). The Yangtze River Delta Integration Plan. *China Briefing*. https://www.china-briefing.com/news/yangtze-river-delta-integration-plan/

Press, K. (2006). A Life Cycle for Clusters?: The Dynamics of Agglomeration, Change, and Adaption. *Springer Science & Business Media.*

Qureshi, I., & Compeau, D. (2009). Assessing between-group differences in information systems research: A comparison of covariance-and component-based SEM. *MIS quarterly*, 197-214.

Richardson, H. W. (1995). Economies and diseconomies of agglomeration. In *Urban agglomeration and economic growth*. Springer, Berlin, Heidelberg. pp. 123-155

Romer, P. M. (1986). Increasing Returns and Long-Run Growth. *Journal of Political Economy.* 94, 1002–1037.

Sadowski, M. M. (2016). Crossroads of The World, Crossroads of The Law: Hong Kong and Macau Legal Systems Approaching 20 Years Post-Transition. *Comp. L. Rev*., 7, 1.

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of psychological research online, 8*(2), 23-74.

Schmidt, A. F., & Finan, C. (2018). Linear regression and the normality assumption. *Journal of clinical epidemiology*, 98, 146-151.

Shalabh, S. (2012). Methods for Estimation and Inference in Modern Econometrics. *Journal of the Royal Statistical Society. Series A (Statistics in Society).* Wiley, *175*(3), 822-840.

Sheng, M., & Gu, C. (2018). Economic growth and development in Macau (1999–2016): The role of the booming gaming industry. *Cities*, 75, 72-80.

Sheng, N., & Tang, U. W. (2013). Zhuhai. *Cities,* 32, 70-79.

Shihe, F. U., Xiaofang, D. O. N. G., & Guojun, C. H. A. I. (2010). Industry specialization, diversification, churning, and unemployment in Chinese cities. *China Economic Review, 21*(4), 508-520.

Smith, A. (2010). The Wealth of Nations: An inquiry into the nature and causes of the Wealth of Nations. *Harriman House Limited.*

Sohlberg, M. (2018). China Free Trade Zones (FTZ): The Definitive Guide for Exporters. *Retrieved from https://www.export2asia.com/blog/china-free-trade-zones/*

Solow, R. M. (1956). A Contribution to the Theory of Economic Growth. *Quarterly Journal of Economics*, 70, 65–94

StataCorp. (2013). STATA Structural Equation Modeling Reference Manual: Release 13. *College Station, TX*: StataCorp LP.

Statistics and Census Service (DSEC). (2019). Macao Statistical Yearbook (various years) [data file]. *Government of Macao Special Administrative Region.* Retrieved on July 29, 2020, from http//dsec.gov.mo.

Storper, M. (1997). The regional world: territorial development in a global economy. *Guilford Press*, New York.

Storper, M. (2004). Territories, flows, and hierarchies in the global economy. *Reading economic geography*, 271.

Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education, 48*(6), 1273-1296.

Tang, J., Williams, A. M., Makkonen, T., & Jiang, J. (2019). Are different types of interfirm linkages conducive to different types of tourism innovation?. *International Journal of Tourism Research, 21*(6), 901-913.

Tarka, P. (2018). An overview of structural equation modeling: its beginnings, historical development, usefulness and controversies in the social sciences. *Quality & quantity, 52*(1), 313-354.

The Heritage Foundation. (2019). *Index of Economic Freedom* (various years) [data file]. *Washington, D.C: Heritage Foundation.* Retrieved July 28, 2020, from heritage.org/index/explore.

Tian, L., Wang, H. H., & Chen, Y. (2010). Spatial externalities in China regional economic growth. *China Economic Review,* 21, S20-S31

UN-Habitat. (2019). The Story of Shenzhen: Its Economic, Social and Environmental Transformation. *United Nations Human Settlements Programme.* 2, 1-172.

Van der Panne, G. (2004). Agglomeration externalities: Marshall versus Jacobs. *Journal of evolutionary economics, 14*(5), 593-604.

Van der Panne, G., & van Beers, C. (2006). On the Marshall–Jacobs controversy: it takes two to tango. *Industrial and Corporate Change, 15*(5), 877-890.

Vogel, R. K., Savitch, H. V., Xu, J., Yeh, A. G., Wu, W., Sancton, A., ... & Shen, J. (2010). Governing global city regions in China and the West. *Progress in Planning, 73*(1), 1-75.

Wang, H., Dong, S., & Mostafizi, A. (2017). Understanding Interdependencies Between Systems Toward Resilient Critical Lifeline Infrastructure in the Pacific Northwest.

Warf, B. (Ed.). (2010). Encyclopedia of geography. *1. Sage.*

Williams, R. (2018, 20 March). *Panel Data 4: Fixed Effects vs Random Effects Models*. Retrieved 25 August, 2020, from https://www3.nd.edu/~rwilliam/stats3/Panel04-FixedVsRandom.pdf

Wong, M. Y., & Kwong, Y. H. (2020). One formula, different trajectories: China's coalition-building and elite dynamics in Hong Kong and Macau. *Critical Asian Studies, 52*(1), 44-66.

World Economic Forum. (2019). *Global Competitiveness Report* (various years) [data file]. *Geneva: World Economic Forum.* Retrieved July 28, 2020, from weforum.org/reports.

Xu, D., Pan, Y., & Beamish, P. W. (2004). The effect of regulative and normative distances on MNE ownership and expatriate strategies. *MIR: Management international review*, 285-307.

Xu, D., & Shenkar, O. (2002). Note: Institutional distance and the multinational enterprise. *Academy of Management review, 27*(4), 608-618.

Yang, C. (2007). Divergent hybrid capitalisms in China: Hong Kong and Taiwanese electronics clusters in Dongguan. *Economic Geography, 83*(4), 395-420.

Yang, J., Xia, X., & Zhang, M. (2018). A Study on Economic Spatial Structure of Urban Agglomerations in Guangdong-Hong Kong-Macao Greater Bay Area. *International Journal of Business and Management, 13*(10).

Yau, C. (2018, November 23. Guangzhou-Shenzhen-Hong Kong Express Rail Link revenues, passengers fall short of estimates two months after launch, MTR figures show. *South China Morning Post.* https://www.scmp.com/news/hong-kong/transport/article/2174782/guangzhou-shenzhen-hong-kong-express-rail-link-revenues

Yuan, K. H., & Bentler, P. M. (2007). Robust procedures in structural equation modeling. In *Handbook of latent variable and related models* (pp. 367-397). North-Holland.

Zhang, H. (2015). How does agglomeration promote the product innovation of Chinese firms?. *China Economic Review,* 35, 105-120.

Zhang, P., & Peeta, S. (2011). A generalized modeling framework to analyze interdependencies among infrastructure systems. *Transportation Research Part B: Methodological, 45*(3), 553-579.