Development in China and Africa

Manuscript committee: Prof. dr. S. Brakman
Prof. dr. E.H. Bulte
Dr. G. Faber
Prof. dr. W.H.J. Hassink
Prof. dr. F.H.M. Verbeeten

This research was funded by the China Scholarship Council in 2010.

ISBN 978-94-91870-11-8
Tjalling C. Koopmans Dissertation Series
USE 022
Printed by Ridderprint, Ridderkerk

Development in China and Africa

Ontwikkeling in China en Afrika
(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor
aan de Universiteit Utrecht
op gezag van de rector magnificus,
prof. dr. G.J. van der Zwaan,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op
vrijdag 31 oktober 2014 des middags te 12.45 uur

door

Shiwei Hu
geboren op 15 mei 1985 te Fushun, China

Promotor:

Prof. dr. C. van Marrewijk

To my parents

致父母

Acknowledgments

It is not easy for me to live and work far away from home and abroad for four years. Looking back on the PhD journey, however, many people helped me, encouraged me and supported me that made the last four years a wonderful experience. Therefore I would like to express my deepest appreciation to many of you without whom this dissertation would not have been possible.

First and foremost, I would like to thank my supervisor, Charles van Marrewijk, for giving me the opportunity to start my research career at school of economics Utrecht University. I still remember how excited I was when I received your offer and how nervous I was when I met you for the first time in 2010. I really appreciate your critical comments on my research, which enable me to understand what the real research looks like and which direction I should work hard on. I really appreciate your encouragement and support on my research, which guide me to finish my PhD on time and give me the confidence to continue my research career in the future. You have always been giving the useful remarks on my work patiently and timely, however leaving a lot of room to think and develop my research independently. It is a great pleasure to work with you and every tiny progress that I have made is indispensible with your guidance. In addition, I am also very grateful of your arrangement for my visit to Xi'an Jiaotong-Liverpool University where I enjoyed every talk we have about Chinese issues during the coffee break. In the end, I look forward to continuing our joint work and being friends in the future.

I thank my reading committee, Prof. dr. Steven Brakman, Prof. dr. Erwin Bulte, Dr. Gerrit Faber, Prof. dr. Wolter Hassink and Prof. dr. Frank Verbeeten for their carefully reading of my dissertation and valuable comments. I am also grateful to my co-author, Prof. dr. Steven Brakman, whose encouragement released my stress in the conference. I thank seminar participants at USE and ETSG 2013 and 2014, and SCERI-CCER 2014 conference for all valuable comments and suggestions given to the chapters in this dissertation.

I am indebted to a number of stuff and colleagues at USE who provide me an accommodating environment in the last four years. I thank the secretaries Mariska, Veronika and Janet. Many thanks to Mariska, from whom I have got continuous help since I applied for the PhD position four years ago. I am also grateful to Mark, Marcel, Joep, Lu, Han-Hsin, Krista and Saara for the thought-provoking academic discussions we had together. I thank my fellow friends, Sarah, Maryam, Bahar, Rebean, Britta, Joyce, Emre, Bastian, Jingyang and Li for the joys and cozy time we had together.

Also I would like to thank all my Chinese friends, Xiaoqian, Can, Xueqing, Yang, Xin, Xu, Jianxi, Yihong and so forth, for sharing this tough and wonderful oversea life together. I cherish every moment we spent together for cooking, birthday and house warming parties, shopping and traveling, which rescued me from the stressful work and homesickness, especially the first year we arrived in Utrecht.

Most importantly, word cannot express my truly gratitude to my parents without whose understanding and supporting I could not get through the four years aboard. 谢谢爸爸妈妈对我多年在外求学工作的理解与支持，对我在海外生活上的关心，对我小脾气的包容。感谢老妈每周末与我闲聊家中琐事，让我缓解工作压力和思乡之苦。感谢老爸每逢在我迷茫困惑的时候及时宽慰我和鼓励我。今年是我离开家乡的第十年，女儿在外父母难免挂心操劳，谢谢你们支持我的每一个决定，谢谢你们总是让我感受到家庭的温暖，无论我身在何方。

Last, I thank my boyfriend, Di, for his love, understanding and tireless support. Thanks to the NAKE course development economics in which we met for the first time. I am indebted to your continuous support both spiritually and academically. You always cheer me up whenever I got frustrated by setbacks in research; you always listen to my questions patiently; you are always happy to share good news with me. For your love and care, I am much grateful than I can tell.

Shiwei Hu
Utrecht, September 2014

Contents

Acknowledgments .... **i**

List of Table .... **vi**

List of Figure .... **vii**

List of Appendix .... **viii**

Chapter 1 Introduction **1**

1.1 The economy of China and Africa **1**

1.2 Places in the literature, questions and contributions **3**

1.2.1 China-Africa relations 3

1.2.2 African export structure 5

1.2.3 China's development 6

1.3 Outline of the thesis **7**

Chapter 2 An empirical analysis of the Africa-China trade puzzle: the role of China's trade policies **9**

2.1 Introduction **9**

2.2 Historical Background and Stylized Facts **10**

2.3 Literature **14**

2.4 Empirical Methodology **18**

2.4.1 The Traditional and Structural Gravity Model 18

2.4.2 Zero-Inflated Poisson Regression 21

2.4.3 The Comparison of Predicted Pattern with the Real World 23

2.4.4 The Modified Model with Trade Policies 24

2.5 Data **27**

2.6 Basic Regression Results for the Standard Gravity Model **29**

2.6.1 Passive Group (first stage, Logit) 31

2.6.2 The Active Group (second stage, Poisson) 31

2.7 The Africa-China Trade Puzzle **32**

2.8 Five Potential Policy Explanations for the Puzzle **34**

2.8.1 Five trade policies 34

2.8.2 Estimating the Policy Impact 37

2.9 Conclusions **40**

2.10 References ... 42
2.11 Appendix ... 45
Chapter 3 Absence of structural change in African trade flows ... 53
3.1 Introduction ... 53
3.2 Related studies ... 55
3.3 Methodology ... 57
3.3.1 Balassa index (BI) ... 57
3.3.2 Harmonic Weighted Mass index (HWM) ... 58
3.3.3 Markov transition probability matrix and mobility index ... 61
3.4 Sample and data sources ... 64
3.5 Results of identifying structural change ... 65
3.5.1 Results with Balassa index ... 65
3.5.2 Results for the HWM index ... 69
3.5.3 Explanations for certain structural breaks ... 73
3.5.4 Markov transition and mobility indices ... 77
3.6 Possible explanations for the absence of structural change ... 82
3.6.1 Persistence in exporting primary and natural-resource intensive goods ... 82
3.6.2 Barriers to shifting to manufacturing production ... 85
3.7 Conclusions ... 90
3.8 References ... 91
3.9 Appendix ... 94
Chapter 4 Regions, agglomerations, and cities in China ... 101
4.1 Introduction ... 101
4.2 Theoretical framework ... 103
4.2.1 Related studies ... 103
4.2.2 Model structure ... 105
4.3 Empirical methodology ... 107
4.3.1 Elasticity test ... 107
4.3.2 Pairwise comparison test ... 108
4.4 Data ... 109
4.4.1 The administrative division of locations ... 109
4.4.2 Skills ... 112
4.4.3 Sectors and occupations ... 113
4.5 Empirical results ... 116
4.5.1 Larger cities are relatively more skilled ... 116
4.5.2 Larger cities specialized in skill-intensive sectors and occupations ... 121

4.6 Conclusions ... 127
4.7 References ... 129
4.8 Appendix ... 131
Chapter 5 Conclusion ... 145
5.1 Main findings ... 145
5.2 policy implications ... 147
5.3 Limitations and future research ... 148
Nederlandse Samenvatting (Dutch Summary) ... 151
Curriculum Vitae ... 153
TKI Dissertation Series ... 155

List of Table

Table 2.1: Basic regression results, Zero-Inflated Poisson regression, 1990-2000 30

Table 2.2: Percent deviation of predicted trade flows 33

Table 2.3: Africa-China trade, 2003-2010 38

Table 3.1: Four Mobility Indices 63

Table 3.2: The highest Balassa index in selected years, 2-digit 66

Table 3.3: Share of sectors with BI>1 in total sectors, percentage (%) 68

Table 3.4:The share of countries with structural changes in each# of years difference (%)71

Table 3.5: All significant structural breaks: transition period, 1-5 years difference 73

Table 3.6: Categories of the Balassa index 78

Table 3.7: Markov 1-year transition matrices for Africa by two categories: the pooled sample 79

Table 3.8: Mobility indices for pooled sample: 1-year transition 79

Table 3.9: Correlation matrix of mobility index 80

Table 4.1: Summary of Chinese administrative division at prefectural and county level . 110

Table 4.2: Population shares of skill group by educational attainment in 2000 and 2010(%) 113

Table 4.3: Average education of employment and population share in each sector 114

Table 4.4: Average education of employment and population share in each occupation.. 115

Table 4.5: Population elasticities of educational groups 117

Table 4.6: Success rate of hypothesis 1 elasticity test: large locations are more skill intensive 119

Table 4.7: Success rate of hypothesis 2 125

List of Figure

Figure 2.1: China's GDP and imports from Africa (1990-2010).......................................12
Figure 2.2: The top 10 African Exporters to China: by Country and Product (1995 & 2005) 13
Figure 2.3: The share of oil imports of China (1995-2010)..14
Figure 2.4: The predicted and actual export flows from Africa to its main trading partners 33
Figure 2.5: China's outward FDI in Africa: stocks and flows (1990-2010)........................34
Figure 3.1: GDP per capita of Africa (1980-2010)...53
Figure 3.2: P-P plot of Balassa index (Angola, 1991-1996)...60
Figure 3.3: The number of African countries with significant structural change...............69
Figure 3.4: The share of the countries with structural change in each sample (%) 70
Figure 3.5: HWM index of Africa by exports concentration...72
Figure 3.6: Natural-resource export concentration: the value and share 75
Figure 3.7: Share of sectors in product group with BI>1, 2-digit.......................................76
Figure 3.8: Correlations of M1-M3 between category I and category II............................80
Figure 3.9: M1-M4 for individual African countries and benchmarks, category I.............81
Figure 3.10: M1-M3 for individual African countries and benchmarks, category II 82
Figure 3.11: Inward FDI stocks and flows in Africa, 1980-2010.......................................83
Figure 3.12: The dynamics of oil imports from Africa and oil prices, 1980-2010.............84
Figure 3.13: Labor force, 1990-2010...86
Figure 3.14: The education of the labor, 1990-2010 87
Figure 3.15 Three variants of concentration, 1980-2010..88
Figure 3.16: Institutional indicators of Africa (average of 1996-2010)...............................89
Figure 4.1: Yancheng prefecture; Jiangsu province, China, 2010.....................................111
Figure 4.2: Population elasticities of skills and years of schooling..................................118
Figure 4.3: Pairwise comparison of six educational attainment levels, subsample...........120
Figure 4.4: Sectors' population elasticities and skill intensities, subsample 122
Figure 4.5: Occupations' population elasticities and skill intensities, subsample............124
Figure 4.6: Pairwise comparison of 19 (15) sectors and 7 occupations, subsample..........126

List of Appendix

Appendix 2.1: List of the African Exporting Countries 45
Appendix 2.2: List of Importing Countries 45
Appendix 2.3: List of the Importing Countries (continued) 46
Appendix 2.4: Variable Descriptions and Sources 47
Appendix 2.5: Variable Descriptions and Sources (continued) 48
Appendix 2.6: China's ETCZs in Africa 49
Appendix 2.7: Descriptive Statistics (Full Sample) 50
Appendix 2.8: Descriptive Statistics (Africa-China Sample) 50
Appendix 2.9: Correlation Matrix for the Full Sample 51
Appendix 2.10: Correlation Matrix for the Africa-China Sample 52
Appendix 3.1: 20 African countries 94
Appendix 3.2: 181 Reference countries 94
Appendix 3.3: Classification of 2-digit sectors 95
Appendix 3.4: Classification of 2-digit sectors (continued) 96
Appendix 3.5: Transition probability matrix for four dropped countries: category I, 1-year 96
Appendix 3.6: Transition probability matrix for individual African countries: 1-year 97
Appendix 3.7: Transition probability matrix for individual African countries: 1-year (continued) 98
Appendix 3.8: Transition probability matrix for individual African countries: 1-year (continued) 99
Appendix 3.9: Mobility indices for individual countries 100
Appendix 4.1: Full name and short name of sector and occupation 131
Appendix 4.2: Population elasticities of educational groups, full sample 132
Appendix 4.3: Population elasticities of skills and years of schooling, full sample 133
Appendix 4.4: Pairwise comparison of six educational attainment, full sample 134
Appendix 4.5: Sectoral employment population elasticities, 2000 135
Appendix 4.6: Sectoral employment population elasticities, 2010 136
Appendix 4.7: Occupational employment population elasticities, 2000 137

Appendix 4.8: Occupational employment population elasticities, 2010 137
Appendix 4.9: Sectors' population elasticities and skill intensities, full sample.............. 138
Appendix 4.10: Occupations' population elasticities and skill intensities, full sample..... 139
Appendix 4.11: Pairwise comparison of 19 sectoral and 7 occupational categories, full sample 140
Appendix 4.12: The pairwise comparison for skills 141
Appendix 4.13: The pairwise comparison for sectors 142
Appendix 4.14: The pairwise comparison for occupations 143

Chapter 1 Introduction

1.1 The economy of China and Africa

Since the beginning of the economic reform in 1978, China's rapid economic growth has impressed the world. Over the last thirty years, the economic reform expanded to various domains in China. Specifically, the initial reform in the agricultural sector strongly increased the productivity in rural areas and made it possible to supply a large number of surplus laborers, who were (partially) implicitly unemployed in the agricultural sector before the reform and now move to the non-agricultural sectors in urban areas. In addition, the economic reform also spread to urban areas. Private sectors and foreign firms started to boom and became increasingly important in urban economy, eroding the dominance of state owned enterprises. All these reforms led to a transition from a central planned economy to a more market-based economy, which in turn benefited economic prosperity in China over time.

In the wave of the globalization, the formerly isolated Chinese economy has become increasingly integrated into the world economy. Through trade flows, financial flows, and foreign aid, China expanded its global presence, especially towards Africa in the past two decades. As the political and business connections between Africa and China are becoming closer and tighter, scholars and policymakers are paying more attention to China's connections with Africa (Broadman, 2007; Wang, 2007; Besada, Wang, and Whalley, 2008; Biggeri, and Sanfilippo, 2009; Cheung, de Haan, Qian, and Yu, 2012). African countries were quite willing to welcome China's changes. We thus observe closer Africa-China relations and surging bilateral trade flows. There is an increasing trend in the share of export values from Africa to China since 1990. The exports to China only accounted for 0.4 percent of African total exports in 1990. This share has increased to 12.9 percent by 2010. Put differently, the value of African exports to China rose from nearly nil in the early 1990s to around 60 billion US dollar in 2010. By 2010, China has become the second largest export market for Africa, ranking second only to the United States, according to the UN Comtrade database.

INTRODUCTION

Meanwhile, African economy achieved good performance in the last two decades with the average annual GDP per capita growth rate of 0.18 percent in the 1990s and 2.26 percent in the 2000s, respectively.[1] This may be lower compared to the East Asia and the Latin America, but it is higher than its negative growth rate in the 1980s (-0.14 percent). The positive economic growth of Africa may be driven by the trade openness, inflow FDI, and improvement in the infrastructure, government, educational environment and so forth. However, Africa still faces a big challenge to sustain its economic growth in the future, as its economy highly relies on the export of primary products. Whether Africa can start economic transformation and upgrade industries are important to the sustainability of the economic growth. Luckily, African countries do have an opportunity to promote their economic transformation. First, African countries extract high resources rents from abundant natural resources in the environment of increasing prices of resources in the international market. The increased resources rents could potentially improve the investment environment and skills of laborers, which in turn contribute to upgrading the industries. Second, along with the global production processes changed over the last twenty years, sectors such as textile and assembling industries start to shift towards Africa thanks to its increasing demography, rapid urbanization and cheap labor costs. Compared to the primary sectors, this shift could be seen as an important lift that may contribute to the industry upgrading and further economic growth in Africa.

Both China and Africa are important players in the world economy today. An analysis of their individual economic activities and the relationships between the two gives better knowledge regarding the development process.

[1] Author's calculation based on African development indicator (ADI) GDP per capita growth (annual %).

1.2 Places in the literature, questions and contributions

1.2.1 China-Africa relations

There was a substantial increase in African exports over the past two decades. In particular, African exports to China have been growing even faster than that to the EU and US. Meanwhile, China has maintained higher rates of economic growth than the EU and US as well. A common view usually attributes the sharp rise of African exports to China to China's fast economic growth. This conjecture is prevailing in the media and has the theoretical foundation based on the standardized international trade model, say gravity model.

Since being proposed by Tinbergen (1962), the gravity model has become a workhorse model in the empirical analysis of international trade.[2] Based on Newton's law of gravitation, this model implies that the attraction between two regions increases in their economic size (usually measured by GDP) and decreases in their bilateral trade cost (usually measured by bilateral distance). This basic version of the model clearly shows that rapid economic growth in one region promotes international trade with another region. The estimated equation fits the data very well, while the fit of the traditional gravity model improves when supplemented with other proxies for trade costs, such as the effect of common language, the common borders and so forth (Anderson, 2010). However, the lack of theoretical foundations makes the gravity model being dubious for trade economists. Several economists proposed the theoretical foundation of the gravity model in different ways, among which, Anderson and van Wincoop (2003) have been viewed as the most important work in the gravity research. The authors not only extended the theoretical foundation of Anderson (1979), but also simplified the model. More importantly, they derived the structural gravity model and stressed the importance of multilateral resistance terms, which was absent in the traditional gravity model and provide a more accurate estimation for subsequence empirical work.

Economic growth is considered as the major driver of the trade development in the analysis of gravity model. However, we find that even if we take into account China's fast

[2] All references are listed at the end of each chapter, except for this chapter.

economic growth, the Africa-China trade is still larger than what the gravity model predicted. In other words, the growth of the Chinese economy relative to the rest of the world cannot fully explain the story of the soaring African export flows to China. We call it the 'Africa-China trade puzzle'. This puzzle indicates a special trade connection between Africa and China that motivates us to find the other drivers of China's rapid imports from Africa.

It is noted that unlike the Africa-EU or the Africa-US trade relationship, the Africa-China trade has certain features that differentiate it from the common trade story. In the China's move to Africa, Chinese government designed a series of trade policies that entailed closer and tighter connections between Africa and China along multiple dimensions for both political and business considerations. These trade policies could be candidates of the hidden drivers for the African exports to China, on top of the factors usually included in the gravity model (Zafar, 2007; Besada et al., 2008; De Grauwe et al., 2012). First of all, China encouraged foreign direct investment (FDI) to Africa, especially to oil, mineral and construction sectors primarily through state owned enterprises. By extracting the natural resources, this so called 'vertical FDI' aims to provide the inputs to the downstream productions for the host country, China. Thus, the vertical FDI from China to Africa accelerates the exports from Africa to China. Second, China is speeding up the pace of trade liberalization by setting up special lists of duty-free import for African goods and establishing a number of the *Economic and Trade Cooperation Zones* in Africa. Both of the activities aim to get rid of trade barriers between China and Africa, which boost the trade flows between the two regions. Third, China's *business and politics separation policy* might explain China's large number of imports from African countries with bad governance. Unlike the EU and US, China has traded with most of the African countries (about 50 countries) since the 1990s. Ignoring the effect of politics both widen and deepen the international trade between China and Africa. Finally, China encourages natural resource imports to satisfy an increasingly huge demand for natural resources in the process of urbanization and industrialization. As a natural resource abundant continent, Africa is a perfect supplier. The resource-seeking trade policies of China further promoted the exports of natural resources from Africa.

Empirical studies of African trade flows remain scare. Besides, current studies either use the gravity model only to explain the trade or focus only on individual trade policy.

Therefore, chapter 2 aims to fill in the gap in the literature by combining both analyses in a comprehensive way. We use a large sample of 20 African countries in a recent period from 1990 to 2010. We adopt an out-of-sample estimation strategy by comparing the predicted export values based on parameters from an estimated gravity model with actual data. The comparison suggests the inability of the gravity model to explain the Africa-China trade relationship. In the end, we supplement with trade policy variables and show that the inclusion of these variables substantially increase our explanatory power.

1.2.2 African export structure

African economic development has received a lot of attention around the world, as on the one hand, Africa is the continent that contains the largest number of poor countries. Therefore its economic growth is important in addressing the problem of poverty, inequality and unemployment for less developed countries. On the other hand, with plentiful endowments of natural resources, Africa would be impetus for growth of the world economy. The question whether African countries can capitalize their resource endowments to maintain their rapid economic growth attracts much attention.

Recovering from the negative economic growth in the 1980s, we do observe positive and increasing economic growth in Africa from 1990 to 2010. It is well recognized that rapid economic growth is usually accompanied by a process of industrial upgrading, namely from primary to manufactory sectors and finally to service sectors. As the change of the export structure to some extent reflects the change of the domestic industrial structure, the rapid economic growth of a given country is usually accompanied by the structural change of export products. For example, having grown rapidly since the 1980s, China saw dramatic changes in its export structure. Manufactured goods have overtaken the place of primary agricultural products as the most important export products. In return, the upgrading structural breaks have a positive effect on economic growth.

Against this background, chapter 3 aims to identify whether there were potential (or missing) similar breaks in African export structure from 1990 to 2010. In principle, there are some forces that engender the change of the international specializations, such as knowledge spillovers and technology transfer (Grossman and Helpman, 1991, Chapter 7). With absence of these drivers, however, the trade pattern is determined by initial conditions and reinforces itself over time (Lucas, 1988; Grossman and Helpman, 1991, Chapter 8). Therefore, we also investigate the potential reasons for the (absence of) structural breaks in African exports over time.

This chapter contributes to the literature in two aspects. First, few studies investigate African trade structure using the latest data. We seek to fill in the gap by analyzing the absence of structural breaks in African export patterns and associated reasons in the last two decades. Second, we utilize a formal statistical analysis to identify dynamics of African trade patterns in the last two decades. Unlike the prior literature (Ariovich, 1979; Ng and Yeats, 2001; Edwards and Schoer, 2002) that relies on partial information of trade patterns, our analysis is based on the entire distribution of revealed comparative advantages, which improves the precision of estimation.

1.2.3 China's development

The distribution of skills and industries across space has been widely discussed in urban economics and new economic geography (NEG). More specifically, several urban studies have identified the distributions of workers with different skills across cities. They find that the workers of higher skills are inclined to stay in larger cities, as they benefit from staying there (Glaeser, 1999; Mori and Turrini, 2005; Bacolod, Blum, and Strange, 2009; Glaeser and Resseger, 2010). The agglomeration of industries, driven by the reallocation of labor and firms, relies on the increasing returns to scale and the decreasing transportation costs. However, Ellison and Glaeser (1997), Duranton and Overman (2005) and Ellison, Glaeser and Kerr (2010) address the significance of agglomeration, localization and coagglomeration of industries, which are relative to the locations of industries being uniformly allocated to the proportion of local population.

Combining these studies, Davis and Dingel (2013) develop a model based on the systems-of-cities, land-use theory and international trade to explore the joint relationship between the skills distribution and the sectoral employment distribution across cities. Unlike the prior studies of Abdel-Rahman and Anas (2004) and Behrens, Duranton and Robert-Nicoud (2012), Davis and Dingel (2013) assume a continuum of skills and indifferent locations within a city. They provide a new perspective to analyze the spatial distribution of skills and sectors. In addition, they also empirically confirm their predictions using US cities' data in 2000.

Based on the theoretical framework of Davis and Dingel (2013), chapter 4 aims to investigate whether larger cities are relatively more skill abundant and specialize relatively more in skill-intensive sectors in China over time. China is a good laboratory for our study for two reasons. First of all, the study of China provides new evidence for less developed countries that might support the theory. Secondly, driven by the economic reform commenced in 1978, China has seen a dramatic transition from a central planned economy to a market-oriented economy. Meanwhile, the distribution of mobile laborers with different skills and industries were more and more market driven. This transition enables us to further investigate the evolution of Chinese cities and the distribution of skills and sectors over time.

1.3 Outline of the thesis

The remaining of this dissertation consists of three empirical chapters. Chapter 2 examines the Africa-China trade from 1990 to 2010. Chapter 3 analyzes the dynamics of African trade patterns in the last two decades and provides potential reasons for the absence of structural change of African exports over time. Chapter 4 provides new evidence for the theory of Davis and Dingel (2013) regarding how the distributions of skills and sectors systematically vary across Chinese cities and evolve over time.

To test the Africa-China trade puzzle, Chapter 2 uses the gravity model to assess the role of China's economic growth in explaining the Africa-China trade relations in the first step. (i) Using the data in 1990-2000, we estimate the general trade pattern, which features the role of output, transportation costs and other institutional factors in determining the trade flows. (ii) We employ the estimated trade pattern with data of explanatory variables in 2001-2010 to predict the trade flows in the same period. By comparing the predicted trade pattern based on the gravity model with the actual trade flows, we confirm the limited role of economic growth in explaining the rise in Africa-China trade flows. We conclude that the classic gravity model fails to fully explain Africa-China trade and therefore justify the puzzle. The second step further explores the hidden factors that have an impact on African exports to China by including China's preferential trade policies. The inclusion of the trade policy variables improves our analysis of the Africa-China trade flows and partly supplements the analysis of gravity model in explaining international trade.

Chapter 3 empirically identifies whether there is a structural break of African exports. We use a sample of 20 African countries from 1990 to 2010. Relying on the revealed comparative advantage, measured by the Balassa index, we identify the export pattern of each Africa country. Then based on the entire distributions of the export patterns, we use two indices to investigate the potential structural breaks in African exports. We find no evidence of structural change in African exports, compared to a benchmark of European countries. This suggests that economic growth is not necessarily related to industrial upgrading, at least in the case of Africa. Moreover, we analyze the potential explanations for the absence of the structural changes in African trade flows. We conclude that the abundant endowments of natural resources, the inadequacy of skilled labor force, certain geographic and institutional factors prevent African exports from upgrading to a manufactures dominant structure.

Chapter 4 explores the spatial distribution of skills and sectors across cities in China. There are two main issues often posed when China took the transition from a central planned economy to a market-oriented economy over the last three decades. First, whether lager cities in China are more skill abundant and more specialized in skill-intensive sectors? Second, whether the distributed pattern of skills and sectors changes during the transition? To answer these two questions, we analyze the data of 6 educational levels, 19 sectors and 7 occupations at three different levels of Chinese cities in 2000 and 2010, respectively. Our results show that the distribution of skills is determined by the cities' size in both years. In particular, larger cities are relatively skill abundant in all three levels of Chinese cities in line with the US patterns. However, unlike the US, the distribution of sectors and occupations across cities shows a pattern of evolution in China over time. More specifically, larger cities produce more highly skill-intensive sectors in 2010 only, but no clear trend in 2000. We attribute this phenomenon to the fast urbanization accompanied with the development of market economies during this period.

Finally, we conclude in chapter 5 and offer policy implications, followed by a discussion of limitations of this research and some suggestions for future studies.

Chapter 2 An empirical analysis of the Africa-China trade puzzle: the role of China's trade policies[3]

2.1 Introduction

Since the start of the Chinese economic reforms in 1978, and especially after the end of the Cold War, China's economy has been growing rapidly. China built ever closer economic ties with the rest of the world, as illustrated by China joining the World Trade Organization (WTO) in 2001. Through trade flows, financial flows, and foreign aid, China expanded its global presence, especially towards Africa in the past two decades (see below). As the political and business connections between Africa and China are becoming closer and tighter, scholars and policymakers are paying more attention to China's connections with Africa (Broadman, 2007; Wang, 2007; Besada, Wang, and Whalley, 2008; Biggeri, and Sanfilippo, 2009; Cheung, de Haan, Qian, and Yu, 2012). African countries were quite willing to welcome China's changes. We thus observe closer Africa-China relations and surging bilateral trade flows. In this chapter we are particularly interested in explaining the Africa-China trade dynamics in the period 1990-2010.

China's fast economic growth since 1990s impresses the world. It is believed to be one of the main reasons for the sharp rise in African exports towards China. Nonetheless, we will argue that China's outstanding economic performance relative to the rest of the world cannot fully explain the striking surge of imports from Africa. The literature regarding Africa–China cooperation hints that the trade policies designed by the Chinese government seem to be the hidden drivers for the remaining surge in trade (Zafar, 2007; Besada et al., 2008; De Grauwe et al., 2012). We thus hypothesize that on top of the China's rapid economic evolution its favorable trade policies also contribute to the surge in African exports to China.

---

[3] This chapter is based on joint research with Charles van Marrewijk.

To identify what we label the 'Africa-China trade puzzle' and to estimate the contribution of preferential trade policies in explaining this puzzle we proceed in three steps, which also summarizes our contribution to the literature. First, we explain the general African trade pattern by estimating a gravity model for a large sample consisting of 28 African exporters and 181 importers in 1990-2000. Second, we compare the actual trade flows in the period 2001-2010 with the predicted export flows based on the first step. Since Africa-China trade rises much faster than the predicted level based on differences in income growth alone we label the extra increase the Africa-China trade puzzle. Third, we include preferential trade policy information to estimate the part of the puzzle that can be explained by policy changes. We will conclude that the policy changes are important but cannot fully explain the Africa-China trade puzzle.

The remainder of this paper is organized as follows. Section 2.2 provides an historical background and the trends in Africa-China trade relation. Section 2.3 reviews the methods and findings in the previous literature. Section 2.4 sets out the empirical methodology employed in this analysis. Section 2.5 presents our data sources and a descriptive summary of the explanatory variables. Section 2.6 presents the estimated trade pattern from a general gravity model using Zero-Inflated Poisson regression. Section 2.7 compares the predicted exports with the actual ones and thus presents the puzzle. Section 2.8 analyzes the contribution of policy variables in a modified gravity model for explaining the puzzle. Finally, section 2.9 concludes the paper.

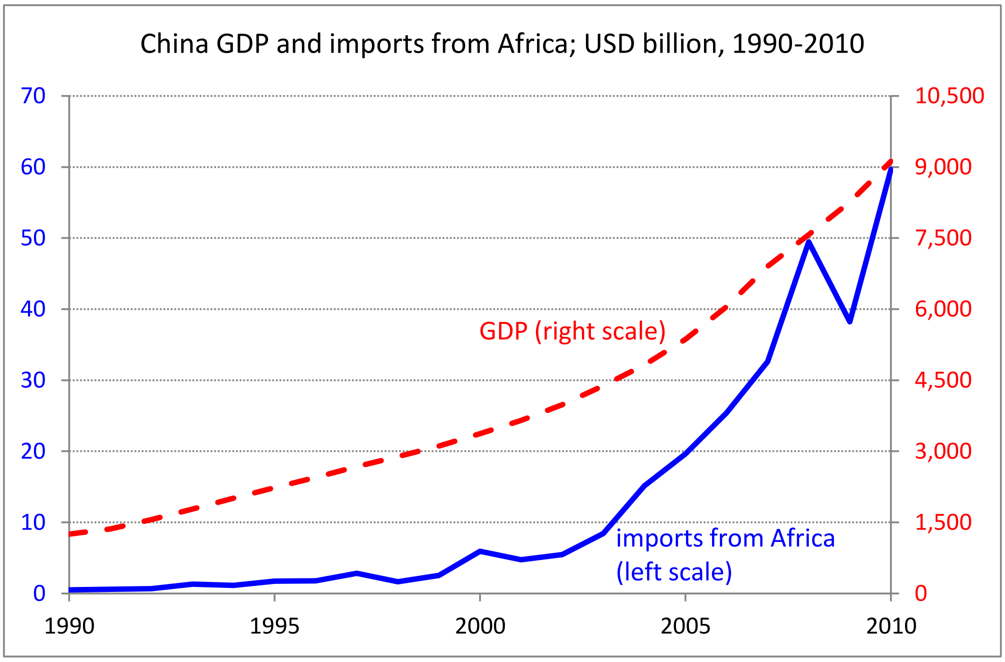
2.2 Historical Background and Stylized Facts

According to the Chinese historical records and archaeological discoveries in Egypt, the trade relations between China and Africa date back at least two thousand years (Gao, 1984). Through the famous "Silk Road", Chinese silk yarn and metal pots travelled across Central Asia and finally arrived in Egypt. Meanwhile, African products, such as elephant tusks, rhinoceros horns and incense, also found their way to China (Gao, 1984; Li, 2009). In the subsequent centuries, along with the development of seafaring techniques and shipbuilding industry, ships were employed in the Africa-China merchant trade. Most Chinese products were shipped to the Middle East and then transported to Eastern Africa (Gao, 1984; Rotberg, 2008). In the Ming dynasty (1368-1644), Chinese ships directly reached Eastern Africa for the first time. The great Chinese navigator, Zheng He, set out three large-scale expeditions to Egypt, Somalia, and Kenya. Gold, silk and porcelain were exchanged with local African products (Rotberg, 2008). Unfortunately, the Africa-China relations cooled down suddenly during the Ching dynasty (A.D.1644-1911), which adopted a closed-door policy with respect to foreign relations. This stopped most of China's trade relations with other continents, including Africa (Gao, 1984). In the final phase of the Ching dynasty and the next half-century China suffered from aggressive wars and civil wars, such that the relations between China and Africa were negligible during that period.

After the foundation of the People's Republic of China (PRC) in 1949, the Africa-China relations improved again. The history of the modern Africa-China interaction can be divided into three periods. The first period starts from 1949 and ends before China's economic reform in 1978. During this post-colonial period China established diplomatic relations with the independent African countries. Although bilateral trade flows were small, China granted unilateral economic assistance to politically intimate African countries in the fields of agriculture and infrastructure (Alden, 2009). Africa-China relations thus basically reflected China's ideological considerations and its foreign policy. The second period (1978–1989) starts from the beginning of China's reform until the end of the Cold War. During this period China dramatically shifted its focus to the domestic economy and cut down the volume of foreign aid. The third period starts in the early 1990s and lasts to the present day. During a visit to Africa in 1996, the Chinese former president Jiang Zemin officially emphasized the importance of economic relations between China and Africa on a non-ideological basis. As a crucial platform for Africa-China cooperation, the Forum on China-Africa Cooperation (henceforth: FOCAC) was established in 2000. It hosts a ministerial conference every three years. In 2006, the Chinese government announced eight policies to support its economic activities with Africa, including encouraging investment, expanding the number of duty-free import products and establishing economic and trade cooperation zones. This is considered a milestone for the Africa-China relations. In the same year, the Chinese government initiated "China's policies for Africa" to highlight the cooperation with African countries in economic fields, especially in trade and investment. All of these policies are expected to expand the breadth and depth of the cooperation between Africa and China. We now briefly review the Africa-China trade relations.

During the past two decades, the remarkable economic growth of China has been widely acknowledged. China's fast development benefited its trading partners as well, see Figure 2.1. On the one hand, China has maintained an annual economic growth rate of 9.9 percent.[4] On the other hand, China has become the second largest export market for Africa. There is an increasing trend in export values from Africa to China since 1990, especially since 2002. In 1990, China only accounted for 0.4 percent of African total exports. By 2010, this share increased to 12.9 percent.[5] The value of African exports to China rose from close to zero in the early 1990s to around 60 billion US dollar in 2010. Though temporarily affected by the economic crisis in 2009, the African exports to China bounced back already in 2010.

Figure 2.1: China's GDP and imports from Africa (1990-2010)



Source: The UN Comtrade database and World Development Indicators 2012 (GDP, PPP (constant 2005 international $)). The African exports to China have been adjusted by the US deflator (base year 2005).

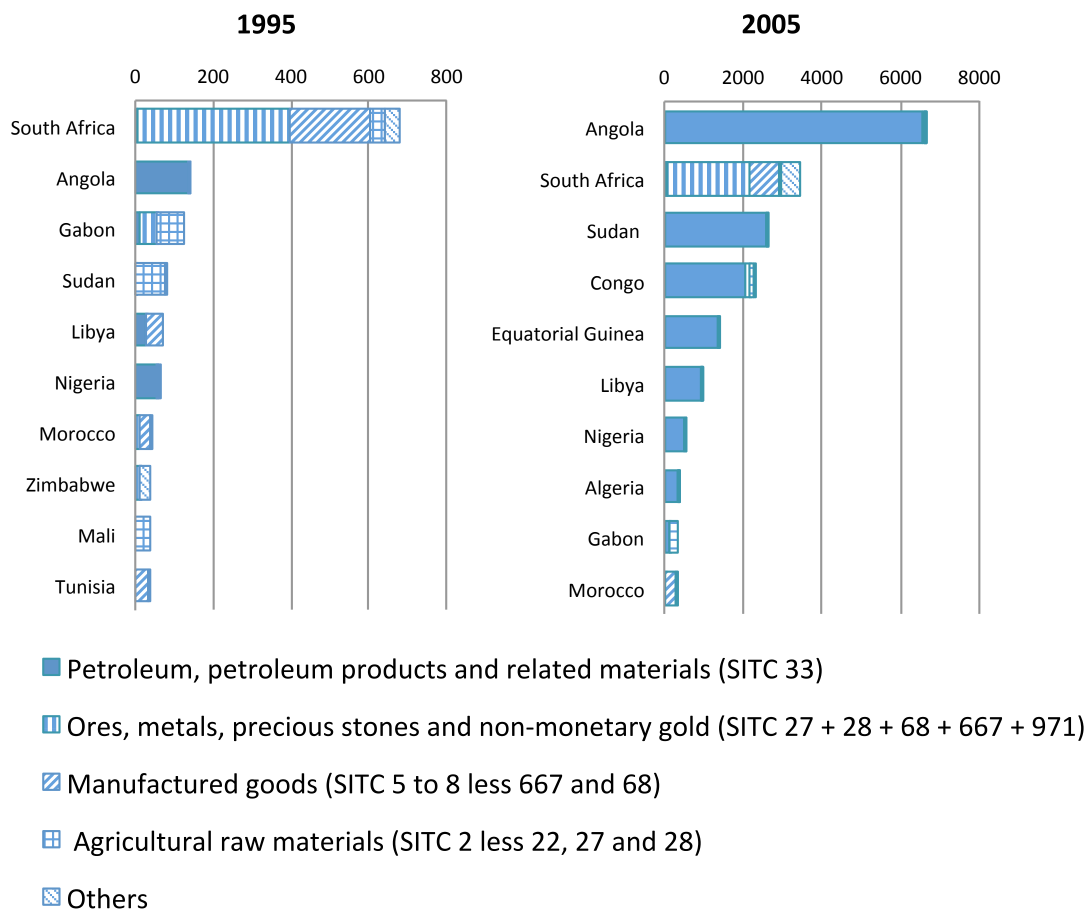
Figure 2.2 decomposes China's main African trading partners by export flows and by products in different periods. Though China has traded with almost all 57 African countries since the 1990s, African exporters are heavily concentrated in a few oil-exporting and mineral-exporting countries. At the beginning of our time period, in 1995, the top 10

[4] The compound annual growth rate = $(GDP_{2010}/GDP_{1990})^{\frac{1}{2010-1990}} - 1$. Source: World Development Indicators 2012 (GDP, PPP (constant 2005 international $)).

[5] Source: The UN Comtrade database.

African exporters accounted for 91 percent of total African exports to China. The largest trading partner, South Africa, accounted for approximately half of the total African exports to China, with metals and manufactured goods as the primary export products. The other main exporters were oil-exporting countries (Angola, Libya and Nigeria) and agricultural producers (Gabon and Sudan). In 2005, the top 10 African exporters still played a decisive role in African exports to China (accounting for about 89 percent of the total). Angola became the largest exporter, with South Africa moving to second place. Oil exporters (Angola, Sudan, Congo, Equatorial Guinea, Libya, Nigeria and Algeria) dominate the top 10 list, since oil accounts for about 70 percent of African exports to China.

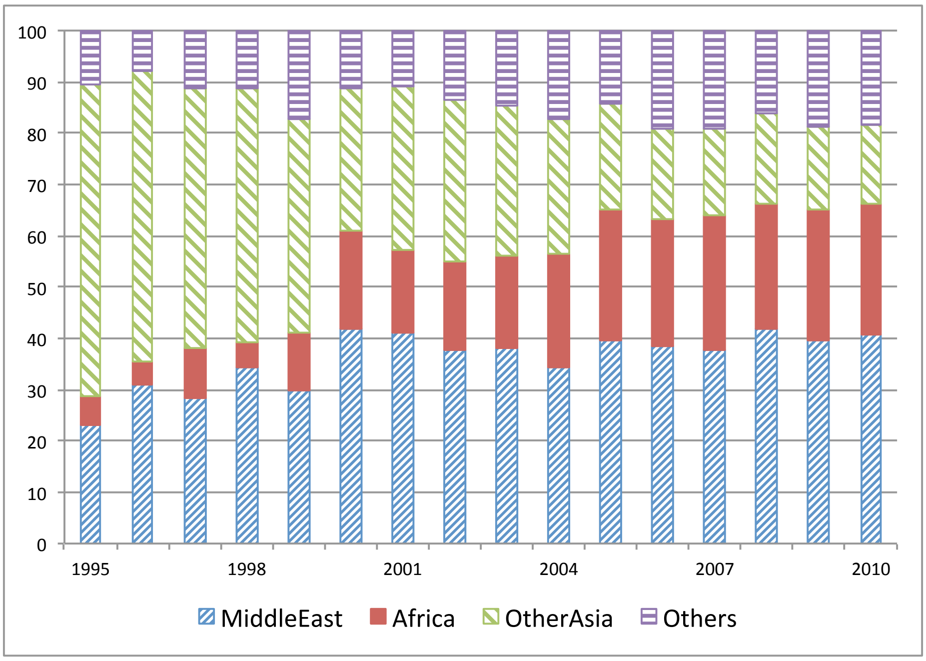
Figure 2.2: The top 10 African Exporters to China: by Country and Product (1995 & 2005)



Source: The UN Comtrade database. The horizontal axis denotes the value of exports in million USD

Figure 2.3 shows the share of different sources of China's oil imports from 1995 to 2010. Initially, 'Other Asia' was the main oil supplier for China, accounting for more than a half of total Chinese oil imports.[6] Since the turning point in the year 2000, the share of oil imports from Other Asia has declined from 28 to 15 percent. By contrast, the share of oil imports from the Middle East to China had increased from 23 percent in 1995 to 42 percent in 2000 and been fluctuating around 40 percent in the past ten years.[7] Moreover, the share of oil imports from Africa increased markedly from 6 percent in 1995 to 26 percent in 2010, making Africa the second largest oil supplier for China in the past decade.

Figure 2.3: The share of oil imports of China (1995-2010)



Source: The UNCTAD database. SITC 33: Petroleum, Petroleum product and related materials.

2.3 Literature

Gravity model has been one of the most empirically successful models in estimating bilateral trade data since Tinbergen (1962). However, the lack of theoretical foundations makes the gravity model being dubious for trade economists. Although Anderson (1979) was the first attempt to theoretically derive the gravity model by specifying the expenditure

---

[6] 'Other Asia' refers to all Asian countries except for China and the Middle East countries.

[7] According to the CIA World Factbook, the Middle East includes 19 countries, namely Armenia, Azerbaijan, Bahrain, Gaza Strip, Georgia, Iran, Iraq, Israel, Jordan, Kuwait, Lebanon, Oman, Qatar, Saudi Arabia, Syria, Turkey, United Arab Emirates, West Bank and Yemen.

function to be a Constant Elasticity of Substitution (CES) function on the demand side structure, his model did not attract much attention among the trade research as his theoretical foundation seemed too complex to be a part of everyday toolkit (Leamer and Levinsohn, 1995). Up until 1995, Trefler (1995) find there are a large number of missing trades in empirical data that is much lower than the prediction of Heckscher-Ohlin-Vanek (HOV) model, while gravity model is one way to measure and explain it. Meanwhile, Leamer and Levinsohn (1995) and Krugman (1995) point out the importance of including bilateral distance in trade studies. The trade economists' attention shifted back to gravity model again. Later on, Eaton and Kortum (2002) proposed an alternative derivation of the theoretical underpinning of gravity model. In contrast with the foundation of Anderson (1979), Eaton and Kortum (2002)'s theory is derived from the supply side structure, which is based on the homogenous goods, iceberg trade costs and Ricardian model. Anderson and van Wincoop (2003) have been viewed as the most important work in the gravity research. The authors not only extended the theoretical foundation of Anderson (1979), but also simplified the model. More importantly, they derived the structural gravity model and stressed the importance of multilateral resistance terms, which was absent in the traditional gravity model and provide a more accurate estimation for subsequence empirical work.

Some studies discuss the Africa-China trade using gravity model. De Grauwe et al. (2012) find that China is unique in trading with African countries from 1996 to 2009 on the basis of the standard gravity model and a set of government indices. Because China is consistently willing to import more from African countries of bad governance compared to other major world economies. Johnston et al. (2014) test whether the African political policies, namely market economy recognition and Taiwan recognition, have positive effects on the African exports to China from 1995 to 2009. They find that the former is associated with increased import from Africa, while the effect of the latter is not statistically significant. Berthelemy (2011) includes the Chinese engagement and the attempt to search for natural resource, captured by the stock of FDI from China and dummy variables for oil rich countries, in the gravity equation from 2004 to 2007. However, these gravity studies did not (or limitedly) stress the special role of Chinese involvement and certain preferential trade policies in facilitating the Africa-China trade in the last ten years, during which the export flows from Africa to China increased even faster than before. Therefore there should be additional factors that may contribute the international trade between Africa and China.

Indeed there are certain studies that discuss the role of China's engagement and trade policies in Africa-China trade, however, none of those depends on the gravity model. China's outward FDI and aid to Africa are the most frequently mentioned. China's outward FDI, trade, and aid are often analyzed in an integral framework (Wang, 2007; Besada et al. 2008), as these three channels are believed to interact with each other (Kaplinsky et al, 2006). Biggeri and Sanllippo (2009) make the first empirical attempt to evaluate the interactions of the above-mentioned channels. They find that China's engagement in Africa is pushed by these interactions and pulled by resources endowment and market potential of the African countries. Cheung et al. (2012) find that China's outward FDI in Africa is affected by trade ties and aid. Broadman (2007) concludes that China's growing trade and outward FDI in Africa are mutually reinforcing. Markusen and Maskus (2002) use a theoretical model to justify that if the origin and destination countries are dissimilar in terms of relative factor endowment and size, then vertical FDI stimulates international trade. FDI from the origin countries can activate some specific under-used comparative advantage in destination countries and in return increase the exports from FDI host countries to origin countries (Zarotiadis, 2008). China and African countries clearly have different country sizes and relative factor endowment. Renard (2011) emphasizes that Chinese outward FDI has helped African countries to build local capacity, transfer technology, and raise exports. Therefore he concludes that large Chinese overseas investments into natural resource sectors created a prospering resource-extracting sector in Africa and promoted resource exports, benefiting both Africa and China.

Zafar (2007) and Besada et al. (2008) assess the effect of China's trade liberalization on its trade relations with Africa. Setting a duty-free import list for the least developed countries and lower average tariffs are expected to promote African exports to China. Minson (2008) finds that China's preferential trade policies tend to facilitate Africa's export capacities, though the effects are modest. Berthelemy (2011) discusses the progress of the Economic and Trade Cooperation Zones (henceforth: ETCZs) in Africa.

The resource-seeking motive for China to involve in Africa has been discussed far and wide. Ademola et al. (2009) argue that China's rapid industrialization increases the demand for natural resources. Besada et al. (2008) add that the stagnant domestic output of natural resources further urges China's move to resource rich African countries. Renard

(2011) concludes that China's imports and outward FDI are highly concentrated and dominated by a few resource-abundant countries.

Some authors are concerned with China's attitude towards the quality of African governance. Anderson and Marcouiller (2002) investigate the relation between trade and quality of governance and argue that high quality of governance would reduce transaction costs and boost trade. On the contrary, China is unique in trading with poorly governed African countries. On top of the gravity study of De Grauwe et al. (2012), Renard (2011) claims that China, unlike Western countries, makes a distinction between business and politics in its engagement in Africa.

In this paper we use gravity model to identify the special trade from Africa to China during the last decade. In particular, the Africa-China trade flows increased faster than that of other territories even taking account of the fast economic growth of China. In the end we propose that the Chinese preferential policies may be the potential reasons.

We use the gravity model in our estimation for two reasons. First, the gravity model has been successfully used in empirical predictions, especially when its theoretical underpinning was found. Leamer and Levinsohn (1995) argue that gravity model produces the clearest and most robust findings in economics. Unlike most international economic models concentrated on two or three-country cases, gravity model represents the economic interaction in a many-country world. In that case, it is convenient to apply the gravity model to the worldwide dataset, which also stimulate the successful use of gravity model in the empirical work. Second, the inclusion of policy impacts into gravity model investigates the efficacy of various policies in promoting trade, which offers a useful tool to policy makers. For this purpose it is appropriate to use the gravity model to estimate the effect of the above-mentioned Chinese preferential policies on promoting the Africa-China trade. As the current gravity studies only take into account the limited factors of Chinese trade policies, our paper try to fill this gap by including more trade policies that may facilitate the Africa-China trade flows.

2.4 Empirical Methodology

In this subsection, we first introduce the traditional and structural gravity model. Second, relying on the structural gravity model, we assess the role of China's economic growth in explaining the increase in Africa-China trade relation. Using the data in the first decade (1990-2000), we estimate the general trade pattern, which features the role of output, transportation costs and other institutional factors in determining the trade flows. Third, we employ the estimated trade pattern with the data in the second decade (2001-2010) to predict the trade flows in the next ten years. We then compare the predicted values with the actual values to see to what extent rising trade flows can be explained by China's (and Africa's) economic growth rates. In the end, we explore the unexplained part of the rise in trade flows (the Africa-China trade puzzle) by analyzing China's preferential trade policies. This only partially explains the additional Africa-China trade flows.

2.4.1 The Traditional and Structural Gravity Model

The traditional gravity model relies on analogy with Newton's law of universal gravitation. The bilateral trade flows (T_{ij}) increase in the economic attribute of the two trading partners (X_i and Y_j) and reduce in the bilateral trade cost (D_{ij}) between the two counties.

$$T_{ij} = X_i^a Y_j^b D_{ij}^c \tag{1}$$

In most applications, it is common to use Gross Domestic Product (GDP) of origin and destination countries as proxies of economic attributes and the bilateral distance as the proxy of the bilateral trade cost. Generally, the estimated coefficients on the economic size GDP cluster close to 1. The estimated equation fits the data very well and usually the model explains 80-90 percent of the variance in bilateral trade flows. While the fit of the traditional gravity model improves when supplemented with other proxies for trade costs, such as the effect of common language, the common borders and so forth (Anderson, 2010).

Although the estimated traditional gravity model has a very good fit to the data, it is not theory-based and biased estimated due to the potential omitted variables. On the basis of the conditions of constant elasticity of substitution (CES) preference and spatial allocation of exporter and importer (Anderson, 1979), and of market clearing for the exporter

(Deardorff, 1998), Anderson and van Wincoop (2003) derive a simplified theoretical foundation of gravity model, namely structural gravity model. In addition, they indicate that the lack of the theoretical underpinning of the traditional gravity model leads to biased estimation due to omitted variables, the multilateral resistance terms. The structural gravity model is given by

$$T_{ij} = X_i^a Y_j^b D_{ij}^c \Pi_i^d P_j^e \tag{2}$$

Where the new derived variables Π_i and P_j are referred as the outward and inward multilateral resistance terms, which are the functions of the bilateral trade cost D_{ij} and the multilateral resistance terms themselves (see Anderson and van Wincoop, 2003, equation (10) and (11)). The intuition of the multilateral resistance terms is that bilateral trade cost alone would not adequately explain the effect of trade frictions on bilateral trade, while bilateral trade also depends on relative trade barriers, which are omitted in the traditional gravity model. Because the trade flow from i to j increases by high trade cost from other origins to j as captured by inward multilateral resistance P_j, and by high resistance on movement to i's alternative destinations as captured by outward multilateral resistance Π_i. In other words, the sale and purchase interact with others and involve all other bilateral trade cost. The multilateral resistance problem is solved with structural gravity equation.

In the empirical estimation method, four major approaches could be used as proxies of the multilateral resistance terms. The first approach is the remoteness index, which calculates each country's average effective distance to or from its partners. However, this approach appears as too weak once the theoretical foundation of gravity became clear. Secondly, Anderson and van Wincoop (2003) derive the important terms 'multilateral resistance' in structural gravity, meanwhile they also propose a method to solve this. By making the additional assumption of symmetric trade cost, they use a contraction-mapping algorithm to calculate the multilateral resistance terms and estimate the equation using a nonlinear least squares. The need of a large amount calculation, however, is the drawback of this method (Anderson, 2010). Third, Hallak (2006) and Romalis (2007) have used ratios and ratios methods, but choosing reference countries makes it difficult to implement. In the end, the preferred method for most empirical trade economists uses fixed effects for the importer and exporter to estimate the multilateral resistance terms in the modern practice. Although estimating gravity model with fixed effects dose not involve strong structural assumptions on the structural gravity, it is easy to implement and gives consistent estimates of the average effects. For this purpose we use fixed effects to measure the effect of multilateral resistance terms. However, the country fixed effect creates too many dummy variables to estimate. Therefore, we use region fixed effect instead. In particular, we group all destination countries into eight regions: EU, US, China, Asia excluding China, South America, Oceania, Africa, and the other countries. As all the origin countries belong to the region of Africa, there is no fixed effect for the exporter.

Many empirical applications assume that both GDP and GDP per capita determine the economic attributes of the exporter and importer. Frankel and Wei (1993) was the first attempt to include the GDP per capita in the gravity studies.[8] They believe that the economic size as captured by GDP has a positive effect on trade, meanwhile the economic development as captured by GDP per capita is also positively related to the trade flows, as the country that becomes more developed tend to specialize more and to trade more. For this purpose equation (2) can be rewritten as:

$$LnT_{ijt} = \beta_1 LnGDP_{it} + \beta_2 LnGDP_{jt} + \beta_3 LnGDP_{pcit} + \beta_4 LnGDP_{pcjt} + \gamma LnDist_{ij} + \alpha' W_{ij} + \phi_j + \varepsilon_{ijt} \quad (3)$$

where LnT_{ijt} denotes the log of export flows from country i to country j in year t; $LnGDP_{it}$ and $LnGDP_{jt}$ are log of GDP of country i and j in year t; $LnGDPpc_{it}$ and $LnGDPpc_{jt}$ denote the log of GDP per capita of exporter and importer in year t. In addition, $LnDist_{ij}$ denotes the log of the time-invariant geographical distance between country i and j, meanwhile W_{ij} is a set of observables, typically used and discussed in the previous gravity regressions, to which bilateral trade barriers are related (Eaton and Kortum, 2002; Anderson and van Wincoop, 2004; Head et al., 2010). Both $LnDist_{ij}$ and W_{ij} measure the bilateral trade cost in structural gravity (2). In this paper, a set of control variables W_{ij} include current or past colonial ties, shared language, shared legal origins, shared a currency union, belonging to the same RTA and shared border. The control variables $W_{ij} = 1$ measures zero trade barriers associated with these variables. On top of these, we also take into account whether the origin and destination countries have access to the sea, which are considered as a part of

[8] They used GNP per capita in their paper.

the bilateral trade cost (see Appendix 2.4 and 2.5 for details.). In the end, the region fixed effect ϕ_j captures the multilateral resistance effect, and ε_{ijt} is an error term.

2.4.2 Zero-Inflated Poisson Regression

When estimating the above model we are confronted with three main econometric challenges: (i) heteroscedasticity in the error term, (ii) the value of zero in the log-linear model, and (iii) the so-called "excess zero" problem. Santos Silva and Tenreyro (2006) emphasize the inconsistency problem in estimating the log-linear form of the gravity model caused by heteroscedasticity in the error term such that its variance depends on other regressors. As a consequence, the ordinary least square (OLS) estimation leads to inconsistent estimates of coefficients of interest.

Santos Silva and Tenreyro (2006) also analyze three possibilities for zero trade values in certain country pairs. First, some country pairs do not trade at all in a given period. For instance, it is not surprising to observe no trade between Chad and the Bahamas in 1990. In this case, the log-linear specification of the gravity model per se is not well-defined. Second, rounding errors might generate zeros. In particular, the value of exports from small or poor countries could be so small that it sometimes cannot reach the minimal value of the unit of measurement. Consequently, these values are automatically and literally recorded as zero in the dataset. These rounding errors depend on country characteristics and lead to inconsistent estimates. Third, zeros might be the result of measurement error. It is possible that when trade data are collected and compiled, some missing values are incorrectly recorded as zeros. This measurement error leads to selection bias and inconsistent estimates.

In addition, the fact that a considerable proportion of country pairs do not trade with one another leads to the "excess zero" problem. Helpman et al. (2008) confirm that half of the country pairs never trade in a set of 158 countries. One common approach in the previous studies is to simply drop zero-trade country pairs from the dataset (Frankel, 1997; Anderson and van Wincoop, 2003). Helpman et al. (2008) argue that disregarding zero-trade country pairs would produce biased estimates and insufficiently exploit the information in the dataset of both trading and non-trading country pairs. Other approaches, for instance, estimating the gravity model using $Ln(T_{ijt} + 1)$ as the dependent variable, or employing Tobit regressions, generally lead to inconsistent estimates of coefficients (Santos Silva and Tenreyro, 2006).

To cope with the zero entries, Razin and Sadka (2007) compare the Heckman selection model with the Tobit model and conclude that the Heckman method is superior to the Tobit and OLS approaches because of its unbiased estimates. In line with the analysis of sample selection, Helpman et al. (2008) propose a two-step estimation procedure. In the first step, they estimate a Probit model which specifies the trade probability as a function of right- hand side variables in the gravity model. Based on the predicted probability, they estimate the log-linear form of the gravity model in the second step. They argue that the procedure corrects for the selection bias and ensures the consistency of parameters of interest. One drawback of this approach is the requirement of an excluded variable for identification in the second stage, which is difficult to find. Moreover, Santos Silva and Tenreyro (2006) argue that NLS estimation can be inefficient in the presence of heteroscedasticity. They prefer the Poisson Pseudo-Maximum-Likelihood estimator (PPML), which is consistent if the conditional mean is in correct specifications. This is not a restrictive method since data need not be Poisson distributed and dependent variables need not be integers to obtain consistent estimates (Gourieroux et al., 1984). Garita and van Marrewijk (2008) and Brakman et al. (2010) use the two-step procedure of Zero-Inflated Negative Binomial regression (ZINB) based on the zero-inflated approach which was originally proposed to handle data with excess zeros by Lambert (1992). This approach assumes two zero-generating processes and two latent groups of observations. An observation in the passive group has the value of zero with probability one while an observation in the active group has a positive probability to have non-zero value. They adopt a logistic regression to model M&A selection in the first stage and a negative binomial regression to model M&A flows in the second stage.

To estimate the general trade pattern in the first decade, we combine the Helpman et al. two-step procedure with Santos Silva and Tenreyro's Poisson regression; we use Zero-Inflated Poisson regression (ZIP) in this paper. Note that Zero-Inflated Negative Binomial regression (ZINB) is an alternative method used in the literature, but this is particularly targeted at count data. The ZIP is applicable for non-count data, such as the trade values analyzed in this paper. Our methodology has the following advantages. First, it provides consistent estimates of the coefficients of interest in the gravity model, correcting for sample selection bias and heteroscedasticity in the trade data. Second, it applies even when the trade data do not follow a Poisson distribution and the dependent variables are positive but non-count data.

The estimation is developed in two stages. In the first step, as the latent group membership is not directly observable; we estimate a Logit regression, see equation (4), which models the group membership as a function of the observed characteristics in equation (3).

$$Passive_{ijt} = \begin{cases} 1 & if \quad T_{ijt} = 0 \\ 0 & if \quad T_{ijt} > 0 \end{cases} \tag{4}$$

where the binary variable $Passive_{ijt}$ is an indicator of membership for country pair ij in the Passive Group in year t ($Passive_{ijt} = 1$) or in the Active Group in year t ($Passive_{ijt} = 0$). By definition, the trade value for a country pair in the Passive Group is always zero. While for a country pair in the Active Group, the trade value is likely to be positive, though the probability of being zero is positive. In the second step, we run a Poisson regression using the PPML estimator as in Santos Silva and Tenreyro (2006) for observations in the active group only. The value of exports has been modeled as a function of the observed characteristics in the first stage, see the standard gravity equation (5) with ϑ_{ijt} as the error term.[9]

$$\begin{aligned} T_{ijt} = exp\ [&\beta_1 LnGDP_{it} + \beta_2 LnGDP_{jt} + \beta_3 LnGDP_{pcit} + \beta_4 LnGDP_{pcjt} + \gamma LnDist_{ij} + \alpha' W_{ij} \\ &+ \phi_j]\vartheta_{ijt},\ if\ Passive = 0 \end{aligned} \tag{5}$$

2.4.3 The Comparison of Predicted Pattern with the Real World

Given the estimated coefficients of the Active Group in equation (5) during 1990-2000, combined with the known value of each explanatory variable during 2001-2010, we calculate the predicted values of African export flows to the major trading partners EU-27, USA, China, and the rest of the world (ROW) in the subsequent ten years, according to the following equation:[10]

[9] Though the observed characteristics in the first and second stage need not be the same, we use the same observed characteristics as in equation (3) for fear of omitted variable bias.

[10] Since 2013 the European Union is composed of 28 sovereign member states. Croatia, which joined that year, is not included in our EU data since this is beyond our data availability.

$$\begin{aligned}\hat{T}_{ijt} &= E(T_{ijt}) \\ &= exp\ [\hat{\beta}_1 LnGDP_{it} + \hat{\beta}_2 LnGDP_{jt} + \hat{\beta}_3 LnGDP_{pcit} + \hat{\beta}_4 LnGDP_{pcjt} + \hat{\gamma}\, LnDist_{ij} + \hat{\alpha}' W_{ij} + \phi_j]\end{aligned} \tag{6}$$

where $\hat{T}_{ijt}$ denotes the projected trade values for country pair *ij* in year *t*, *t* = 2001,…, 2010 for all explanatory variables, and $\hat{\beta}_1$ to $\hat{\alpha}$ are the estimated coefficients of the Active Group in the first decade.[11] Availability of the values of export flows in the second decade enables us to further compare the projected export values with the actual ones by calculating the percent deviation, defined as:

$$\begin{aligned}Percent\ deviation &= \frac{T_{ijt} - \hat{T}_{ijt}}{\hat{T}_{ijt}} \times 100\% \\ &= \frac{Actual\ exports - Predicted\ exports}{Predicted\ exports} \times 100\%\end{aligned} \tag{7}$$

where *t* = 2001,…, 2010. This term assesses the predicting power of the standard gravity model and the explanatory power of China's economic development relative to the rest of the world. A significant deviation from the forecasted trade pattern indicates that other factors may play an important role in the rise of Africa-China trade flows.

2.4.4 The Modified Model with Trade Policies

Since we do find a significant underestimation of Africa-China's trade pattern compared with other trade relations, we try to better understand this Africa-China trade puzzle by integrating China's specific trade policies with African resources. More specifically, we assess the role of five trade policies.[12] First, China's outward FDI in Africa is expected to have an influence on African exports to China. Second, China is speeding up the pace of trade liberalization by getting rid of trade barriers. We are concerned with the impact of the duty-free list for promoting African exports. Third, the Economic and Trade Cooperation Zones are established to expand trade flows. Fourth, China's *business and politics*

---

[11] A rationale to adopt a Poisson regression instead of the OLS or the log-linear specification of the gravity equation is to circumvent the problem of Jensen's inequality, which indicates $E(lny \neq lnE(y))$. Put differently, Poisson regressions guarantee the consistency of the expectation of the predicted value of export flows.

[12] Though aid is treated as an important factor in explaining trade, we cannot incorporate aid in our empirical analysis, restricted by lack of China's aid data.

separation policy might explain China's large number of imports from African countries with bad governance. Fifth, we take into account China's resource-seeking motive embedded in trade policies.

We estimate the following equation that combines output and trade policies in an integral framework:[13]

$$\begin{aligned} T_{ijt} = exp\ [&\eta_1 LnGDP_{it} + \eta_2 LnGDP_{jt} + \eta_3 LnDist_{ij} + \eta_4 Landlocked_i + \delta_1 LnFDI_{i(t-1)} + \delta_2 DF_{it} \\ &+ \delta_3 Detcz_{it} + \delta_4 RL_{it} + \delta_5 CC_{it} + \delta_6 Doil_{it} + \delta_7 Dmineral_{it} + \phi_i]\zeta_{ijt} \end{aligned} \tag{8}$$

where i refers to African country i and j refers to the export destination (China) for t = 2003, · · · , 2010.[14] The individual unobserved effect for African country i is ϕ_i, the vector $\eta = [\eta_1, \ldots, \eta_4]$ measures the effects of the standard gravity variables, and the vector $\delta = [\delta_1, \ldots, \delta_7]$ evaluates the effects of the added policy variables.

During the period of 2003-2010, China traded with every African country in our sample. Therefore, the two-stage estimation procedure is not necessary and we only estimate the second step. Because the White test shows that heteroscedasticity is still present, we prefer a Poisson regression. To control for unobserved heterogeneity, we employ a fixed effect Poisson regression. The fixed effect estimator manages the unobserved heterogeneity that is allowed to be correlated with all regressors.[15] To avoid the problem of endogeneity, we lag the FDI variable by one period.

Compared with the standard gravity model in section 2.4.2, we omit several dummy variables, because China shares no colonial ties, no common languages, no legal origins, no common currencies, no regional trade agreements, and no common borders with African countries. Besides, China has access to seas, so the landlocked dummy for China becomes degenerate. Moreover, since China is the unique export destination, we observe a high correlation between China's GDP and its GDP per capita, such that we omit China's GDP per capita (see correlation Appendix 2.10). For symmetry, we omit GDP per capita of

[13] See Appendix 2.5 for details.
[14] Limited by data availability for China's outward FDI, we only estimate for the period 2003-2010.
[15] In section 2.4.1, regional fixed effects capture some unobserved heterogeneity.

African countries as well. In the end, we retain $LnGDP_{it}$ and $LnGDP_{jt}$, which measures the economic size of the trading partners, and $LnDist_{ij}$ and $Landlocked_i$, which captures transportation costs.

The newly added trade policy variables are as follows: $LnFDI_{i(t-1)}$ is the logarithm of the stocks of Chinese outward foreign direct investment in African country *i*. DF_{it} is the share of the value for duty-free goods in total exports from African country *i* to China in year *t*. This indicator captures the effect of diminishing trade barriers. $Detcz_{it}$ is a dummy variable taking a value of 1 when China establishes an Economic and Trade Cooperation Zone (ETCZ) in country *i* in year *t* and zero otherwise. We take into account the quality of governance of African countries. RL_{it} is the indicator of rule of law for country *i* in year *t* and CC_{it} is the indicator of control of corruption for country *i* in year *t* (see section 2.5 for detailed information). Regarding the resource-seeking motive, we incorporate African resource endowments. $Doil_{it}$ is a dummy variable taking a value of 1 if country *i* is abundant in oil resources. $Dmineral_{it}$ is a dummy variable taking a value of 1 if country *i* is abundant in mineral resources.

In order to evaluate the contribution of the newly included trade policy variables, we also estimate the corresponding gravity model for Africa-China trade relation as follows:

$$T_{ijt} = exp\ [\eta_1 LnGDP_{it} + \eta_2 LnGDP_{jt} + \eta_3 LnDist_{ij} + \eta_4 Landlocked_i + \phi_i]\upsilon_{ijt} \tag{9}$$

Since equation (8) and equation (9) are nested models, we are able to compare Pseudo R^2s and conduct a likelihood ratio test. Alternatively, to correct the effect of GDP, we re-estimate a corrected version of equation (8) as follows:

$$\begin{aligned}\tilde{T}_{ijt} = exp\ [&\eta_3 LnDist_{ij} + \eta_4 Landlocked_i + \delta_1 LnFDI_{i(t-1)} + \delta_2 DF_{it} + \delta_3 Detcz_{it} \\ &+ \delta_4 RL_{it} + \delta_5 CC_{it} + \delta_6 Doil_{it} + \delta_7 Dmineral_{it} + \phi_i]\varsigma_{ijt}\end{aligned} \tag{10}$$

where $\tilde{T}_{ijt} = T_{ijt} - \hat{\beta}_1 LnGDP_{it} - \hat{\beta}_2 LnGDP_{jt}$, $\hat{\beta}_1$ and $\hat{\beta}_2$ are elasticities of trade to output estimated in equation (5). To take one step further, we correct all variables in the gravity equation (5) by subtracting them from export values. We re-estimate the effects of the newly added trade policies on the corrected trade values as a robustness check. The model is as follows:

$$\begin{aligned}\breve{T}_{ijt} = exp\ [\delta_1 LnFDI_{i(t-1)} + \delta_2 DF_{it} + \delta_3 Detcz_{it} + \delta_4 RL_{it} + \delta_5 CC_{it} + \delta_6 Doil_{it} \\ + \delta_7 Dmineral_{it} + \phi_i]\psi_{ijt}\end{aligned} \tag{11}$$

where $\breve{T}_{ijt} = T_{ijt} - \hat{\beta}_1 LnGDP_{it} - \hat{\beta}_2 LnGDP_{jt} - \hat{\gamma} LnDist_{ij} - \hat{\alpha} Landlocked_i$, $\hat{\beta}_1$ to $\hat{\alpha}$ are estimated coefficients in equation (5).

2.5 Data

Our sample for the general trade pattern covers 28 African exporters and 181 worldwide importers.[16] Our data set thus consists of 105,840 observations of bilateral export flows (28 × 180 country pairs and 21 years). The list of the African exporting countries is reported in Appendix 2.1 and the list of importing countries is reported in Appendix 2.2 and Appendix 2.3. Data on African exports come from the UN Comtrade Database, adjusted by the US GDP deflator (base year 2005) from the World Development Indicators (WDI, 2012). Data on GDP PPP (constant 2005 international dollars) and GDP per capita PPP (constant 2005 international dollars) come from the World Bank's (2012) World Development Indicators. Data on distance and dummies indicating contiguity, common language and colonial ties are constructed from the CEPII database-GeoDist. The dummies for common currency union and common legal origin are from the CEPII database-Gravity dataset (2006). The dummy for sharing the same Regional Trade Agreement (RTA) is constructed from the CEPII database-Gravity dataset (2006), complemented with data from the WTO.[17] The information on landlocked countries is constructed from the CIA's World Factbook.[18]

---

[16] For country lists see Appendix 2.1, Appendix 2.2 and Appendix 2.3. There are 57 African countries in 2010 (United Nations, 2011). However, not every country is important in African exports. Hence, we set requirements for our sample as follows. First, the country should have a minimum population of one million in 2010. Second, we select countries with more than 10 years annual data, which are well distributed during the period 1990 to 2010. At the same time, the countries' total exports values should rank in the top 20 in Africa in 1990, 2000 and 2010, respectively. In the end, 28 African countries are included in this sample. The export value of the 28 selected African countries accounted for 88 percent of the total exports of 57 African countries in 2010.

[17] http://rtais.wto.org/UI/PublicAllRTAList.aspx

[18] http://www.cia.gov/library/publications/the-world-factbook/fields/2060.html#ay

The sample for analyzing Africa-China trade is also constructed from multiple data sources. Unfortunately, there is no existing database for relevant Chinese data. In most cases, we have to collect data from the government bulletins and other informal data sources.[19]

The official data of China's outward FDI that comply with the IMF standard are available only after 2003. We obtain China's outward FDI stocks in Africa from the 2010 Statistical Bulletin of China's Outward Foreign Direct Investment, issued by the Ministry of Commerce. Since the depreciation issues of FDI stocks have been taken into account when compiled, we simply adjust the data with the US GDP deflator (base year 2005) from the World Development Indicators (WDI, 2012).

The information regarding China's list of duty-free import products is constructed from the bulletins published by the General Administration of Customs.[20] Combined with HS 8-digit data from the China Customs Statistics Yearbook, we further calculate the share of the value for duty-free products in total exports for particular African countries to which the duty-free policy applies. Due to data collection problems, we lack the information in year 2009.

The dummies of Economic and Trade Cooperation Zones (ETCZs) are constructed from the official websites for the Ministry of Commerce (MOFCOM) and those for each Economic and Trade Cooperation Zone individually.[21]

We use the quality of governance indicators from the Worldwide Governance Indicators database, conducted by Kaufmann, Kraay and Mastruzzi (2010). They report estimated values for six indices indicating six aspects of the quality of governance for each country. Each index ranges from -2.5 to 2.5, where a higher value indicates better governance. In this paper, the control of corruption and the rule of law, which have a fundamental influence on international trade, are adopted as proxies for governance quality.

---

[19] Although the data is collected from various sources, it is the best available public data.
[20] Source: General Administration of Customs of the People's Republic of China, website http://www.customs.gov.cn/publish/portal0/. The duty-free policy started since 2005 for certain sectors in the least developed countries.
[21] See Appendix 2.6 for detailed information.

The information on resource abundance for each African country is compiled from the UNCTAD database. We adopt the criteria of resource abundance as being a major player in the global export market of certain natural resource. In other words, we rule out the resource-rich countries with little resource exports restricted by limited production capacity or large domestic resource consumption. Thus, a country is treated as being abundant in oil or mineral resources if it ranks top 50 worldwide in terms of net exports in a given year. In particular, 'oil' refers to petroleum, petroleum products and related materials (SITC 33) and 'mineral' refers to ores and metals (SITC 27+28+68).

Descriptive statistics and correlation matrices are provided in Appendices 2.7-2.10.

2.6 Basic Regression Results for the Standard Gravity Model

Table 2.1 presents the estimation results for two basic specifications of the standard gravity model.[22] Basic I estimates equation (3) without region fixed effects while Basic II includes these fixed effects. The results are substantially unchanged when including the region fixed effects in terms of signs and significance, except that common legal origin and sharing common border no longer determine the probability of an observation being in the passive group. Since region fixed effects are widely adopted in previous studies, we restrict our discussion below to the estimates of the Basic II specification.

In general, the dummies indicating landlocked importers and RTA determine whether country pairs ever trade but have no influence on the volume of trade. In contrast, colonial ties, currency union, common border and legal origin, which have significant effects on the volume of African exports, play no role in determining the probability of bilateral trade.

[22] There are totally 105,840 observations of the whole period 1990-2010. However the regression in Table 2.1 only includes the time period from 1990 to 2000 with 55,440 observations (28×180×11years). During that period, there were some missing values of real GDP in both origin and destination countries causing by various reasons, e.g. civil war, hyperinflation and so on. The final sample consists of 49,519 observations, including 21,971 non-zero observations.

Table 2.1: Basic regression results, Zero-Inflated Poisson regression, 1990-2000

	(1)	(2)	(3)	(4)	(5)
A. Passive group, Logit.	Basic I		Basic II		Std Dev
$Ln(GDP_{origin})$	-0.326***	[-33]	-0.354***	[-36]	1.24
$Ln(GDP_{destination})$	-0.650***	[-77]	-0.662***	[-78]	2.26
$Ln(GDPpc_{origin})$	0.057***	[6]	0.065***	[6]	0.95
$Ln(GDPpc_{destination})$	-0.293***	[-31]	-0.276***	[-30]	1.28
$Ln(Dist)$	0.445***	[36]	0.057**	[4]	0.69
Colonial Ties	-0.269		0.034		0.08
Language	-0.946***	[-61]	-0.874***	[-58]	0.42
Legal	0.046*	[5]	0.008		0.49
Currency Union	-0.076		-0.136		0.13
RTA	-1.332***	[-74]	-1.371***	[-75]	0.22
Common Border	0.263***	[30]	-0.018		0.15
$Landlocked_{origin}$	0.151***	[16]	0.152***	[16]	0.41
$Landlocked_{destination}$	0.345***	[41]	0.397***	[49]	0.41
B. Active group, Poisson.	Basic I		Basic II		Std
$Ln(GDP_{origin})$	0.722***	[145]	0.734***	[148]	1.24
$Ln(GDP_{destination})$	0.868***	[611]	0.897***	[659]	2.26
$Ln(GDPpc_{origin})$	0.302***	[47]	0.308***	[26]	0.95
$Ln(GDPpc_{destination})$	0.302***	[33]	0.183***	[34]	1.28
$Ln(Dist)$	-0.523***	[-30]	-0.331***	[-20]	0.69
Colonial Ties	0.521***	[68]	0.200**	[22]	0.08
Language	0.381***	[46]	0.523***	[69]	0.42
Legal	0.334***	[40]	0.330***	[39]	0.49
Currency Union	2.205***	[807]	2.134***	[745]	0.13
RTA	0.023		-0.040		0.22
Common Border	0.567***	[76]	0.767***	[115]	0.15
$Landlocked_{origin}$	-0.535***	[-41]	-0.590***	[-45]	0.41
$Landlocked_{destination}$	-0.046		0.017		0.41
Observations	49,519		49,519		
Nonzero obs.	21,971		21,971		
Region fixed effects	No		Yes		
Wald Chi	7029.11		8165.81		

Notes: Dependent variable is the value of African export flows (constant 2005 international $). ***, ** and * denote significance at 1, 5 and 10 percent level, respectively. Column (1) and (3) show the estimated coefficients. For variables significant at least 10 percent level, the coefficients could be interpreted as elasticities. Columns (2) and (4) show the incidence rate ratios in brackets, which indicate the percent change in odds ratio (passive group) or in the expected values of trade flows (active group) if a regressor alters by one standard deviation, ceteris paribus. The incidence rate ratio (IRR) is calculated as follows: for continuous variables, IRR = 100 × [exp(coefficient × Std Dev) − 1]; while for dummy variables, IRR = 100 × [exp(coefficient) − 1] due to a discrete change from 0 to 1. Column (5) shows the standard deviation (= Std Dev) of each variable.

To interpret the economic effect of the estimated coefficients, we use the odds ratio and incidence rate ratio (IRR) as in Garita and van Marrewijk (2008). In the Logit model, the odds ratio measures the probability of an observation being in the passive group relative to the probability of being in the active group. If b denotes the estimated coefficient, and σ

denotes a standard deviation for continuous variables and a unit change for dummy variables, respectively, then the *odds ratio* = $e^{b\sigma}$, indicating the odds of being in the Passive Group versus that of being in the active one in the Logit model.[23] It measures a change in odds of being in the passive group corresponding to a σ-size shock to certain variable, ceteris paribus. In the second step, we report $IRR = 100 \times (e^{b\sigma} - 1)$, which measures the percent change in the expected value for a σ-size shock to a variable, holding other regressors constant.

2.6.1 Passive Group (first stage, Logit)

The estimation results for the first stage Logit regression are presented in Panel A. Most regressors are statistically significant and important in determining the membership of the Passive Group as well as the Active Group, with the exception of colonial ties, legal origin, currency union and contiguity.

Generally speaking, larger economies as measured by larger GDP are less likely to be in the Passive Group. This effect is particularly large for importers, for which a standard deviation size reduction in importer GDP increases the probability of falling in the Passive Group by 78 percent. In other words, the likelihood of bilateral trade between the two countries would decrease considerably. As for economic development, the signs of coefficients of GDP per capita for exporters and importers are opposite. It might be surprising at first glance that economic development imposes a positive effect on the likelihood of being in the Passive Group. Note, however, that the economic magnitude of exporter GDP per capita is so small that it is almost negligible. As for geographic factors, like distance and landlocked conditions, we find significant effects, as expected. In particular, countries far away or having no access to sea are less likely to trade. Finally, countries sharing common language and Regional Trade Agreement are less likely to belong to the Passive Group.

2.6.2 The Active Group (second stage, Poisson)

We are more interested the estimation results of the second stage Poisson regression for the Active Group presented in Panel B. Most regressors are statistically significant and

[23] In Table 2.1, we report $100 \times (e^{b\sigma} - 1)$, that is the percent change in odds.

important in determining the value of export trade, with the exception of the landlocked condition of importers and RTA.

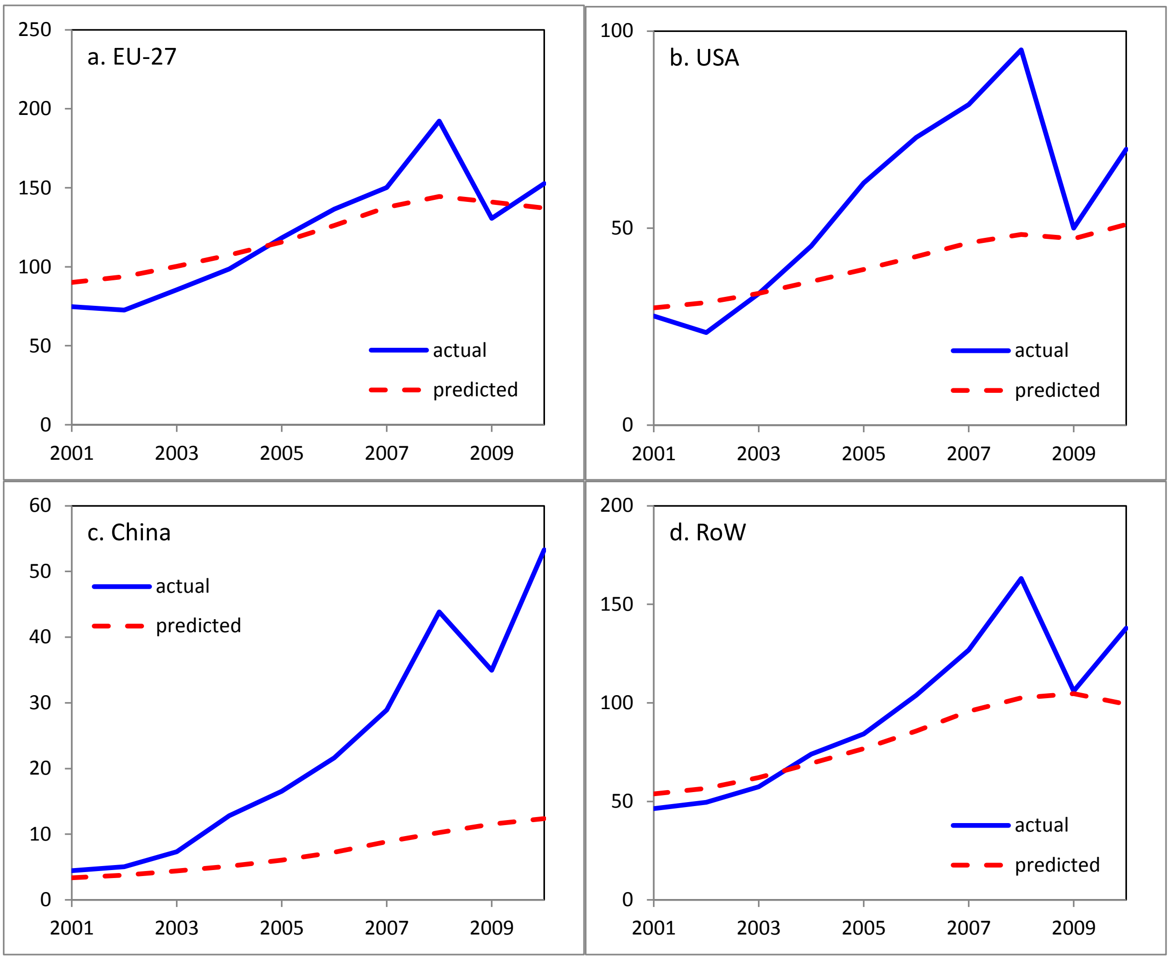
Both GDP and GDP per capita for both exporters and importers have positive and statistically significant effects on the value of exports. If, for example, importer GDP increases by one standard deviation, then the volume of exports increases by 659 percent. We thus expect that importing countries involved in a process of rapid economic growth have rapidly rising export flows from African countries. This certainly holds for China as we discussed above. As for the geographical variables, the distance between trading partners and being landlocked reduce trade flows values, while sharing a common border has the opposite effect. Trading partners with previous or current colonial ties or with a common language or legal origin trade more, as expected. Finally, being within a currency union generates a substantial, positive effect on bilateral trade flows: country-pairs that share the same currency union have 745 percent higher trade values than those using different currencies.

2.7 The Africa-China Trade Puzzle

To test the explanation and predicting power of the gravity model, we use the estimated trade pattern in the first decade of our data set (the coefficients of the Active Group) to predict the value of export flows in the second decade. We then compare the actual values with the predicted ones.

Figure 2.4 graphically presents the predicted export flows (dashed line) and the actual export flows (solid line) for African exporters with their major trading partners, EU-27, USA, China, and the rest of the world (ROW). The USA seems to maintain a pretty stable position in this period. The share of EU-27 in African exports has been declining since the 1990s (not shown). On the contrary, China is becoming more important as a major trading partner of Africa. Although the general trend is rising trade flows, there is a sharp drop in 2009 as a result of the Great Recession, followed by a sharp recovery in 2010. As discussed below, the Africa-China trade relation is consistently underestimated.

Figure 2.4: The predicted and actual export flows from Africa to its main trading partners



Source: author calculations; RoW = Rest of World.

Table 2.2: Percent deviation of predicted trade flows

Year	EU-27	USA	China	ROW
2001	-16.9	-6.8	32.3	-13.9
2002	-22.5	-24.5	33.5	-12.3
2003	-14.7	-0.2	66.3	-7.5
2004	-8.2	24.7	150.5	6.8
2005	2.4	55.5	173.7	9.8
2006	8.1	70.7	197.4	21.2
2007	9.1	75.8	226.0	32.4
2008	33.1	96.7	326.7	59.1
2009	-7.4	5.7	203.4	1.3
2010	11.4	37.4	330.1	38.9

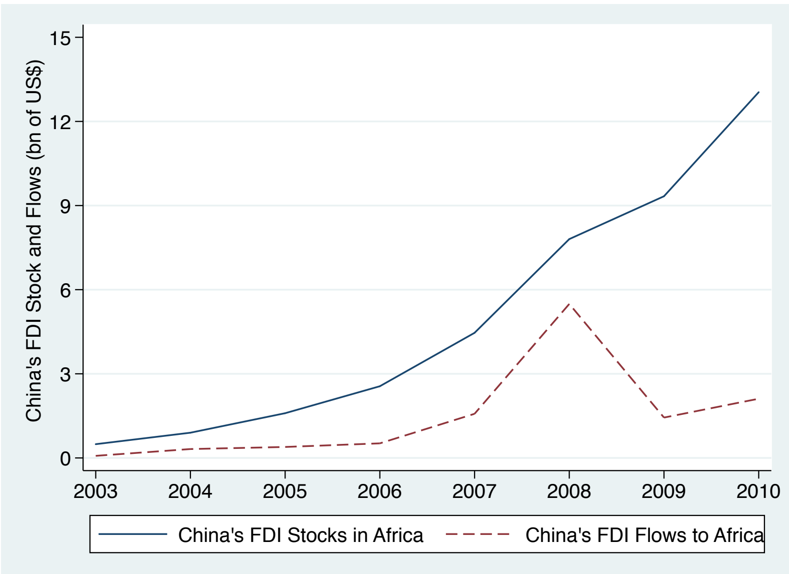
To quantify the deviation from the model prediction we report the percent deviations in Table 2.2. We find initial overestimations and subsequent, mild underestimations for EU-27, USA and the rest of the world. When comparing China with other African trading partners, we find that Africa-China trade flows have been consistently underestimated and increasingly so as time passes by. Taking the deviations for the other countries/regions as a benchmark, China's underestimation is exceptional. We even find a deviation of 330 percent in 2010. The explanatory power of the general gravity model and the rapid economic growth of China relative to the rest of the world is thus not sufficient to explain the rising Africa-China trade flows, hence the 'Africa-China trade puzzle'.

2.8 Five Potential Policy Explanations for the Puzzle

2.8.1 Five trade policies

What could potentially explain the Africa-China trade puzzle? We need to identify other changes during the decade not discussed and incorporated so far in the analysis. As already discussed above, China actively changed several trade policies relative to African countries in this period. We first briefly review the main policy changes and then estimate their potential contribution for explaining the Africa-China trade puzzle in the next sub-section. We focus on 5 policies, namely Foreign Direct Investment, Targeted Trade Liberalization, Economic & Trade Cooperation Zones, No Political Requirements, and the Search for Raw Materials.

Figure 2.5: China's outward FDI in Africa: stocks and flows (1990-2010)



Source: The 2010 Statistical Bulletin of China's Outward Foreign Direct Investment.

1. *Foreign Direct Investment*. Stimulated by the "Going Global" strategy proposed in 2002, China has accelerated its overseas investment. In particular, Chinese investment to Africa has been booming since 2000, especially since 2006 (see Figure 2.5). Though Africa accounts for a small proportion (4.1% in 2010) of all of China's outward FDI, its rapid growth is worth noting. In 2010, Africa attracted 13 billion US dollar of China's outward FDI stocks, which was 26.5 times that of 2003. From 2003 to 2010, Chinese FDI stocks in Africa rose by 60 percent annually.[24] Although the FDI flows declined after the 2008 financial crisis, the FDI stocks still increased, only less rapidly. By 2010, China had invested in 50 African countries, covering 85 percent of all countries on the continent. The primary destinations of China's outward FDI include South Africa, Nigeria, Zambia, Algeria, the Democratic Republic of Congo, Sudan, Niger, Ethiopia, Angola and Egypt. Together, these countries absorbed 76 percent of the FDI stocks in Africa in 2010. A distinct feature of the China's investment in Africa is the vital roles of the state-owned enterprises (SOE) and the state-owned financial institutions. Kaplinsky and Morris (2009) point out that most of the Chinese firms investing in oil, mineral and construction sectors are state-owned. These firms obtain subsidies or low-cost loans from the state-own banks, such as China Export-Import Bank and China Development Bank.

2. *Targeted Trade Liberalization*. China has been accelerating the pace of trade liberalization over the past twenty years. The turning point is the entry to the WTO in 2001, forcing China to get rid of a range of trade barriers. To increase market access for African countries, the Chinese government started setting up a list of duty-free African exports (including 190 products in 8-digit sectors in HS1996)[25] for certain least-developed African countries since 2005. The duty-free list further expanded to cover 444 import sectors in 2007 and 4762 sectors in 2010. By the end of 2010, the duty-free products cover all 2-digit sectors of HS1996 and accounted for 60 percent of all African export sectors. In addition, the Chinese government commited to expand the duty-free list to 95 percent of all African export products in the near future. This duty-exemption policy is expected to contribute to the rise of African exports to China, thereby stimulating the development of the African economy.

[24] The compound annual growth rate = $(\frac{FDI_{2010}}{FDI_{2003}})^{\frac{1}{2010-2003}} - 1$

[25] Source: General Administration of Customs of the People's Republic of China.

3. *Economic and Trade Cooperation Zones*. During the Beijing summit of FOCAC (Forum on China-Africa Cooperation) in 2006, the Chinese government committed to strengthening economic cooperation with Africa by establishing several Economic and Trade Cooperation Zones (ETCZs) in Africa. The objective of the ETCZs is to improve production capacity (particularly export potential) by agglomerating sectors with positive externalities. Firms in the ETCZs are entitled to tax-exemption, favorably-priced land and duty-waivers. Up to 2010 five ETCZs have been constructed, located in Zambia, Nigeria (two ETCZs), Ethiopia, and Egypt. Additional ETCZs in Mauritius, Algeria and Botswana are discussed (see Appendix 2.6 for more details).

4. *No Political Requirements*. China has followed a consistent strategy of separating business decisions from political decisions in Africa that is to do business without any political restrictions. Unlike most Western countries, which set political requirements for trade and aid flows with African countries, China offers unconditional aid to them. Likewise, China has been seeking to build closer economic ties with African countries isolated from the Western world because of their bad governance, such as Sudan and Zimbabwe. As a consequence, China is now the largest buyer of Sudan's crude oil.

5. *Search for Raw Materials*. China's involvement in Africa also reflects its quest for raw materials, especially natural resources. By 2005 China has become the second largest oil consumer in the world.[26] China has also become highly dependent on foreign mineral supplies in view of the declining domestic mineral output. These developments combined with the instable geopolitical situation in some parts of the world explain why China is in search of a reliable supply of natural resources from various locations. Since Africa is the last continent where the main oil supplies and mineral reserves are not dominated by Western companies, the choice of diversification towards Africa is understandable. As a consequence, China has been investing billions of dollars in African resource sectors (see point 1 above) and has been importing large amounts of crude oil, metals, timber, and cotton from Africa.

---

[26] Source: NationMaster.com

2.8.2 Estimating the Policy Impact

Table 2.3 presents the results of estimating the impact of various policy changes on Africa-China trade flows from 2003 to 2010. Column (1) presents an unrestricted baseline estimation based on equation (8), with the corresponding incidence rate ratios (IRRs) reported in column (2). Note that the estimated impact of both origin and destination GDP on the trade flows is higher than the comparable benchmark numbers reported in Table 2.1 (see Basic II with regional fixed effects). This is, of course, not surprising in view of the consistent underestimation of the actual trade flows using this benchmark, as discussed above. For comparison purposes, column (3) provides a similar unrestricted baseline estimation without the five policy changes discussed above. Note that in this case the estimated impact of China's GDP is even higher while that for the origin GDP is lower. It is fair to argue that column (1) underestimates the contribution of the five policies discussed above for explaining the Africa-China trade puzzle in view of the additional explanatory power attributed to origin and destination GDP compared to the benchmark case. To correct for this discrepancy, columns (4) and (5) estimate the impact of the policies after correcting the trade flows for the benchmark GDP influence reported in Table 2.1 (in addition correcting for distance and being landlocked in column (5). We discuss the estimated impact of the five policies in turn below.

1. *Foreign Direct Investment*. The impact of China's rising FDI into Africa is estimated to increase trade flows, that is FDI and trade are complements rather than substitutes (see also Biggeri and Sanfilippo, 2009). The estimated coefficient is statistically significant, but small and even declines if we correct trade flows for the benchmark GDP effects. The *economic* impact of FDI on trade flows as measured by the incidence rate ratio is therefore very limited. It might be argued, on the other hand, that a considerable proportion of China's outward FDI flows to African infrastructure and resource sectors, which spills over to other sectors and eventually enhances aggregate local production capacity in African recipients. FDI may thus have a more indirect effect.

Table 2.3: Africa-China trade, 2003-2010[27]

	(1)	(2)	(3)	(4)	(5)
Dependent variables	T		T	$\tilde{T}$	$\check{T}$
$Ln(GDP_{origin})_t$	0.914***	[149]	0.356***		
	(0.000)		(0.000)		
$Ln(GDP_{China})_t$	1.415***	[312]	1.862***		
	(0.000)		(0.000)		
$Ln(FDI_{origin})_{(t-1)}$	0.012***	[1]		0.077***	0.077***
	(0.000)			(0.017)	(0.018)
$Duty\ Free_{origin(t)}$	-0.019***	[-2]		-0.135	-0.049
	(0.000)			(0.407)	(0.455)
$Detcz_{origin(t)}$	0.425***	[53]		0.924***	0.971***
	(0.000)			(0.230)	(0.210)
$Rule\ of\ Law_{origin(t)}$	-0.911***	[-60]		-0.839*	-0.709**
	(0.000)			(0.340)	(0.421)
$Control\ of\ Corruption_{origin(t)}$	-0.640***	[-48]		-1.287***	-1.359***
	(0.000)			(0.446)	(0.483)
$Doil_{origin(t)}$	0.329***	[39]		0.406**	0.420**
	(0.000)			(0.189)	(0.191)
$Dmineral_{origin(t)}$	0.558***	[75]		0.571***	0.585***
	(0.000)			(0.097)	(0.121)
Observations	174		188	174	174
Number of country pairs	27		27	27	27
McFadden R^2	0.039		0.020	0.020	0.021

Notes: In columns (1) and (3) the dependent variable is the value of African exports (in constant 2005 international $). Column (2) reports the IRR in brackets. In columns (4) and (5) the dependent variable is $\tilde{T}$ and $\check{T}$, see below; ***, ** and * denote significance at the 1, 5 and 10 percent level, respectively. We drop Zimbabwe because of a lack of reliable data for its real GDP, due to hyperinflation since 2000.

$\tilde{T} = T - 0.734 \times Ln(GDPorigin) - 0.897 \times Ln(GDPChina)$

$\check{T} = T - 0.734 \times Ln(GDPorigin) - 0.897 \times Ln(GDPChina) + 0.331 \times Ln(Dist_{ij}) + 0.590 \times Landlockedorigin.$

2. *Targeted Trade Liberalization*. The impact of trade liberalization on the size of trade flows as measured by the coefficient on duty-free products is, rather surprisingly at first sight, negative rather than positive. Please note, however, that this policy is specifically targeting certain sectors of the *least developed* African countries. Some important exporters are excluded from this favorable policy, such as South Africa, Libya, Algeria and Nigeria. In addition, some major components of China's imports, such as oil, timber and cotton, were not on the duty-free product list from 2005 to 2009. Oil and timber were only added in July 2010, while cotton was still excluded. In that sense, the estimated effect is

[27] Since we have an unbalanced panel and missing values for some variables, the observation is not 216 (27 countries×8 years).

not too surprising. More importantly, the economic impact of the unrestricted estimation is very small, while the effect becomes insignificant once we correct trade flows for the benchmark GDP effects.

3. *Economic and Trade Cooperation Zones*. Although they have been set up only recently, the estimated impact of the Economic and Trade Cooperation Zones (ETCZs) on the size of trade flows is positive and significant. For the first time the economic impact measured by the incidence rate ratio is also important: the estimated trade flows increase by 53 percent. In addition, the estimated coefficient more than doubles once we correct trade flows for the benchmark GDP effects.

4. *No Political Requirements*. Two variables are indicative of the impact of separating political from business decisions, namely the impact of the rule of law and the control for corruption. Consistent with De Grauwe et al. (2012), we find that China trades more with countries with a bad rule of law and with corrupt governments, in contrast to Western countries, which have a preference for trading with African countries with better governance. The EU, for example, has strict requirements on the governance quality of trading partners. By rewarding countries that meet the conditions and leaving out others, they create the possibility for China to fill the void and expand trade flows with these isolated countries. Similarly, Cheung et al. (2012) find that law and order risk and corruption risk in Africa tends to encourage Chinese investments. An alternative explanation is the fact that the majority of Chinese trading entities are state-owned enterprises which are less risk-averse since they are supported by the Chinese government under adverse special circumstances. In any case, the estimated economic impact of both the rule of law and corruption as measured by the incidence rate ratio is large at 60 percent and 48 percent, respectively. Moreover, the effect is about the same for rule of law and strengthened for corruption once we correct trade flows for the benchmark GDP effects.

5. *Search for Raw Materials*. The search for raw materials is measured separately for oil and for minerals. In both cases the estimated coefficient is positive and significant, indicating that the resource-seeking motive indeed plays a role for China in expanding trade flows with Africa, see also De Grauwe et al. (2012). The economic impact of the search for raw materials, measured by the incidence rate ratio, is sizeable in both cases,

namely 39 percent for oil and 75 percent for minerals. In both cases the estimated coefficients rise if we correct trade flows for the benchmark GDP effects.

If we take the reported McFadden (pseudo) R^2 as an indicator of explanatory power, then the share of the Africa-China trade puzzle explained by the five policies investigated above is limited in all cases (columns 1, 4, and 5). Unfortunately, these numbers cannot be interpreted independently (Cameron and Trivedi, 2005). We therefore cannot directly evaluate the explanatory power of the model, let alone the contribution of the newly added policy variables. What we can do, however, is to evaluate these numbers in a nested setting. To this end column (3) presents an unrestricted estimation without the policy variables. As we can see in Table 2.3, the pseudo R^2 declines relative to column (1). A formal likelihood ratio test of the two nested models concludes that column (1) is preferred, such that the policy variables are jointly significant.[28]

2.9 Conclusions

We investigate the dynamics and determinants of African exports to China in the past two decades. Along with the rapid growth of the Chinese economy we observe a rapid expansion of China's imports from Africa, such that China is now Africa's second largest export market. We argue that the relative economic expansion of China cannot fully explain the rising trade flows with Africa. To clarify, we first estimate a standard gravity model of Africa's trade flows with the rest of the world, making full use of the recent theoretical and empirical developments in this area, up to 2000. We then compare the predicted post-2000 trade flows with the actual flows to conclude that the standard model consistently underestimated Africa-China trade flows even if we take China's economic development into full consideration. We label this phenomenon the Africa-China trade puzzle.

[28] The total explanatory power measured this way is only about 0.02, either by looking at columns (4) or (5) or by taking the difference between columns (3) and (1). One of the concerns is that the use of value added data rather than trade data may ameliorate the puzzle. However, this concern is not relevant in our context as gross exports to China are very close to value added to China because most exports concentrate on natural resources. The other caveat of our analysis is GDP is too crude as trade with China fits the needs of China perfectly because of the sector composition. Ideally, sector demand variables are better measure than GDP, however, the data of sector demand variables are not readily available.

Second, we investigate the extent to which five different Chinese post-2000 policies, particularly relative to Africa, can explain the Africa-China trade puzzle. The policies investigated are: Foreign Direct Investment, Targeted Trade Liberalization, Economic & Trade Cooperation Zones, No Political Requirements, and the Search for Raw Materials. We find that Targeted Trade Liberalization had no impact on trade flows while Foreign Direct Investment had a significantly positive, but economically negligible impact on those flows. In contrast, the other three policies *do* have an economically sizeable impact. More specifically, trade flows with Economic & Trade Cooperation Zones are substantially larger, as are trade flows focused on the Search for Raw Materials. Moreover, China fills the void left by other countries by trading substantially more with countries with weak rule of law and high levels of corruption (No Political Requirements). Taken together, the five policies above have a statistically significant contribution in explaining the Africa-China trade puzzle, but nonetheless only explain a small fraction of this puzzle.

2.10 References

Ademola, O., A. Bankole, and A. Adewuyi (2009). China-Africa Trade Relations: Insight from AERC Scoping Studies. *European Journal of Development Research*, 21, 485–505.

Alden, C. (2009). China in Africa. *Survival*, 47(3), 147–164.

Anderson, J. E. (1979). A Theoretical Foundation for the Gravity Equation. *American Economic Review*, 63, 106–116.

Anderson, J. E., and D. Macrouiller (2002). Insecurity and the Pattern of Trade: An Empirical Investigation. *The Review of Economics and Statistics*, 84(2), 342–352.

Anderson, J. E., and van Wincoop, E. (2003). Gravity with Gravitas: A Solution to the Border Puzzle. *American Economic Review*, 93, 170–192.

Anderson, J. E., and van Wincoop, E. (2004). Trade Costs. *Journal of Economic Literature*, *42*, 691-751.

Anderson, J. E (2010). The gravity model. *NBER working paper* no. w16576, National Bureau of Economic Research.

Berthelemy, Jean- Claude (2011). China's Engagement and Aid Effectiveness in Africa, Series no. 129, African Development Bank, Tunis, Tunisia.

Besada, H., Y. Wang, and J. Whalley (2008). China's Growing Economic Activity in Africa. *NBER working paper* no. 14024, The National Bureau of Economic Research.

Biggeri, M., and M. Sanfilippo (2009). Understanding China's Move into Africa: an Empirical Analysis. *Journal of Chinese Economic and Business Studies*, 7(1), 31–54.

Brakman, S., G. Garita, H. Garretsen, and C. van Marrewijk (2010). Economic and Financial Integration and the Rise if Cross-Border M&As: The Gravity Model in International Trade. Cambridge University Press, Cambridge.

Broadman, H. (2007). *Africa's Slik Road–China and India's New Economic Frontier*. World Bank, Washington, D.C.

Cameron, C., and P. Trivedi (2005). *Microeconometrics: Methods and Applications*. Cambridge University Press, New York.

Cheung, Y., J. de Haan, X. Qian, and S. Yu (2012). China's Outward Direct Investment in Africa. *Review of International Economics*, 20(2), 201–220.

De Grauwe, P., R. Houssa, and G. Piccillo (2012). African Trade Dynamics: Is China a Different Trading Partner. *Journal of Chinese Economic and Business Studies*, 10(1), 15–45.

Deardorff, A. (1998). Determinants of Bilateral Trade: Does Gravity Work in a Neoclassical World? in Jeffrey A. Frankel (ed), The Regionalization of the World Economy. University of Chicago Press, Chicago.

Eaton, J., and S. Kortum (2002). Technology, Geography and Trade. *Econometrica*, 70(5), 1741–1779.

Frenkel, J. and Wei, S.-J. (1993). Trade Blocs and Currency Blocs. *NBER Working Paper* no. 4335, The National Bureau of Economic Research.

Frankel, J. (1997). *Regional Trading Blocks in the World Economic System*. Institute of International Economics, Washington, D.C.

Gao, J. (1984). China and Africa: The Development of Relations over Many Centuries. *African Affairs*, 83(331), 241–250.

Garita, G., and C. van Marrewijk (2008). Countries of a Feather Flock Together-Mergers and Acquisitions in the Global Economy. Mimeo, Tinbergen Institute.

Gourieroux, C., A. MMonfort, and A. Trognon (1984). Pseudo Maximum Likelihood Methods: Applications to Poisson Models. *Econometrica*, 52, 701–720.

Hallak, J. C. (2006). Product quality and the direction of trade. *Journal of International Economics*, *68*(1), 238-265.

Head, K., T. Mayer, and J. Ries (2010). The Erosion of Colonial Trade Linkages after Independence. *Journal of International Economics*, 81, 1–14.

Helpmam, E., M. Melitz, and Y. Rubinstein (2008). Estimating Trade Flows: Trading Partners and Trading Volumes. *Quaterly Journal of Economics*, 123, 441–487.

Johnston, L. A., Morgan, S. L., and Wang, Y. (2014). The Gravity of China's African Export Promise. *The World Economy*.

Kaplinsky, R., D. McCormick, and M. Morris (2006). The Impact of China on Sub Saharan Africa. Mimeo, Institute of Development Studies, University of Sussex.

Kaplinsky, R., and M. Morris (2009). Chinese FDI in SubSaharan Africa: Engaging with Large Dragons. *The European Journal of Development Research*, 21(4), 551–569.

Kaufmann, D., A. Kraay, and M. Mastruzzi (2010). The Worldwide Governance Indicators: A Summary of Methodology, Data and Analytical Issues. Working paper no. 5430, World Bank Policy Research.

Krugman, P. (1995). Increasing returns, imperfect competition and the positive theory of international trade. *Handbook of international economics*,*3*, 1243-1277.

Lambert, D. (1992). Zero-Inflated Poisson Regression, With an Application to Defects in Manufacturing. *Technometrics*, 34(1), 1–14.

Leamer, E. E., and Levinsohn, J. (1995). International trade theory: the evidence. *Handbook of international economics*, *3*, 1339-1394.

Li, A. (2009). Review of 30 Years of Sino-African Relations. *West Asia and Africa*, 4.

Markusen, J., and K. Maskus (2002). Discriminating among Alternative Theories of the Multinational Enterprise. *Review of International Economics*, 10(4), 694–707.

Minson, A. (2008). China’s Preferential Trade Policy for Africa: China in Africa. Mimeo, South Africa Insittute of International Affairs.

Rasin, A., and E. Sadka (2007). *Foreign Direct Investment: Analysis of Aggregate Flows.* Princeton University Press, Princeton.

Renard, Mary-Francoise (2011). China's Trade and FDI in Africa, Series no. 126, African Development Bank, Tunis, Tunisia.

Romalis, J. (2007). NAFTA's and CUSFTA's Impact on International Trade. *The Review of Economics and Statistics*, *89*(3), 416-435.

Rotberg, R. (2008). *China into Africa: Trade, Aid and Influence*. Brookings Institution Press, Washington.

Santos Silva, J., and S. Tenreyro (2006). The Log of Gravity. *Review of Economics and Statistics*, 88(4), 641–658.

Tinbergen, J. (1962). Shaping the World Economy: Suggestions for an International Economic Policy. The Twentieth Century Fund, New York.

Trefler, D. (1995). The case of the missing trade and other mysteries*. The American Economic Review*, 1029-1046.

Wang, J. (2007). What Drives China's Growing Role in Africa. International monetary fund working paper, International Monetary Fund.

Zafar, A. (2007). The growing relationship between China and Sub-Saharan Africa: Macroeconomic, trade, investment, and aid links. *The World Bank Research Observer*, 22(1), 103-130.

Zarotiadis, G. (2008). FDI and international trade relations: a theoretical approach. *International trade and finance association working papers*, International Trade and Finance Association.

2.11 Appendix

Appendix 2.1: List of the African Exporting Countries

Algeria	Libya	South Africa
Angola	Madagascar	Sudan
Cameroon	Malawi	Tanzania
Cote d'Ivoire	Mali	Togo
Egypt, Arab Rep.	Mauritania	Tunisia
Ethiopia	Mauritius	Uganda
Gabon	Morocco	Zambia
Ghana	Mozambique	Zimbabwe
Guinea	Nigeria	
Kenya	Senegal	

Appendix 2.2: List of Importing Countries

Afghanistan	Ghana	Oman
Albania	Greece	Pakistan
Algeria	Grenada	Palau
Angola	Guatemala	Panama
Antigua and Barbuda Guinea	Papua	New Guinea
Argentina	Guinea-Bissau	Paraguay
Armenia	Guyana	Peru
Australia	Haiti	Philippines
Austria	Honduras	Poland
Azerbaijan	Hong Kong SAR, China	Portugal
Bahamas, The	Hungary	Qatar
Bahrain	Iceland	Romania
Bangladesh	India	Russian Federation
Barbados	Indonesia	Rwanda
Belarus	Iran, Islamic Rep.	Samoa
Belgium	Iraq	Sao Tome and Principe
Belize	Ireland	Saudi Arabia
Benin	Israel	Senegal
Bhutan	Italy	Seychelles
Bolivia	Jamaica	Sierra Leone
Bosnia and Herzegovina	Japan	Singapore
Botswana	Jordan	Slovak Republic
Brazil	Kazakhstan	Slovenia
Brunei Darussalam	Kenya	Solomon Islands
Bulgaria	Kiribati	South Africa
Burkina Faso	Korea, Rep.	Spain
Burundi	Kuwait	Sri Lanka
Cambodia	Kyrgyz Republic	St. Kitts and Nevis
Cameroon	Lao PDR	St. Lucia

Appendix 2.3: List of the Importing Countries (continued)

Canada	Latvia	St. Vincent and the Grenadines
Cape Verde	Lebanon	Sudan
Central African Republic	Lesotho	Suriname
Chad	Liberia	Swaziland
Chile	Libya	Sweden
China	Lithuania	Switzerland
Colombia	Luxembourg	Syrian Arab Republic
Comoros	Macao SAR, China	Tajikistan
Congo, Rep.	Macedonia, FYR	Tanzania
Costa Rica	Madagascar	Thailand
Cote d'Ivoire	Malawi	Timor-Leste
Croatia	Malaysia	Togo
Cyprus	Maldives	Tonga
Czech Republic	Mali	Trinidad and Tobago
Denmark	Malta	Tunisia
Djibouti	Mauritania	Turkey
Dominica	Mauritius	Turkmenistan
Dominican Republic	Mexico	Uganda
Ecuador	Micronesia, Fed. Sts.	Ukraine
Egypt, Arab Rep.	Moldova	United Arab Emirates
El Salvador	Mongolia	United Kingdom
Equatorial Guinea	Morocco	United States
Eritrea	Mozambique	Uruguay
Estonia	Myanmar	Uzbekistan
Ethiopia	Namibia	Vanuatu
Fiji	Nepal	Venezuela, RB
Finland	Netherlands	Vietnam
France	New Zealand	Yemen, Rep.
Gabon	Nicaragua	Zambia
Gambia, The	Niger	Zimbabwe
Georgia	Nigeria	
Germany	Norway	

Appendix 2.4: Variable Descriptions and Sources

Variable	Description	Sources
$Exports_{ijt}$	The volume of exports flow from exporter *i* to importer *j*, adjusted by US GDP deflator (base year 2005).	UN Comtrade Database.
$GDP_{origin(t)}$	Gross domestic products PPP (constant 2005 international $) of the exporting country *i* (in logs).	The World Bank's (2012) World Development Indicators
$GDP_{destination(t)}$	Gross domestic products PPP (constant 2005 international $) of the importing country *j* (in logs).	The World Bank's (2012) World Development Indicators
$GDP_{pc_{origin(t)}}$	Gross domestic products per capita PPP (constant 2005 international $) of the exporting country *i* (in logs).	The World Bank's (2012) World Development Indicators
$GDP_{pc_{destination(t)}}$	Gross domestic products per capita PPP (constant 2005 international $) of the importing country *j* (in logs).	The World Bank's (2012) World Development Indicators
$Dist$	The distance (in km) between exporter *i*'s and importer *j*'s capitals (in logs).	The CEPII database-GeoDist.
$Common\ Border$	A binary variables equals one if exporter *i* and importer *j* are neighbors that meet a common physical boundary, and zero otherwise.	The CEPII database-GeoDist.
$Landlocked_{origin}$	A binary variables equals one if exporter *i* has no coastline or direct access to sea, and zero otherwise.	The CIA's World Factbook .
$Landlocked_{destination}$	A binary variables equals one if importer *j* has no coastline or direct access to sea, and zero otherwise.	The CIA's World Factbook .
$Colonial\ \ Ties$	A binary variables equals one if exporter *i* ever colonized importer *j* or vice versa, and zero otherwise.	The CEPII database-GeoDist.
$Language$	A binary variables equals one if exporter *i* and importer *j* share a common language, and zero otherwise.	The CEPII database-GeoDist.
$Legal$	A binary variables equals one if exporter *i* and importer *j* share a common legal origin, and zero otherwise.	The CEPII database-Gravity dataset (2006).

Appendix 2.5: Variable Descriptions and Sources (continued)

Variable	Description	Sources
Currency Union	A binary variables equals one if exporter *i* and importer *j* use the same currency or if within the country pair money was interchangeable at 1 : 1 exchange rate for an extended period of time, and zero otherwise.	The CEPII database-Gravity dataset (2006).
RTA	A binary variables equals one if exporter *i* and importer *j* belong to a common regional trade agreement, and zero otherwise.	The CEPII database-Gravity dataset (2006) and World Trade Organization (WTO).
$Ln(FDI)_{origin(t-1)}$	China's outward Foreign Direct Investment (FDI) in African country *i*, adjusted by US GDP deflator (base year 2005), in logs, lagged by one period.	The 2010 Statistical Bulletin of China's Outward Foreign Direct Investment.
$Duty\ Free_{origin(t)}$	The share of the value for duty-free export products in total export flows for country *i*.	The China Customs Statistics Yearbook and the bulletins published by the General Administration of Customs of the PRC.
$Detcz_{origin(t)}$	A binary variables equals one if China establishes an Economic and Trade Cooperation Zone (ETCZ) in country *i*.	The official websites for the Ministry of Commence of the PRC and those for each ETCZ individually.
$Rule\ of\ Law_{origin(t)}$	The governance indicator measuring the level of rule of law in country *i*. It ranges from -2.5 to 2.5, where a higher value indicates better governance.	The World Governance Indicators database.
$Control\ of\ Corruption_{origin(t)}$	The governance indicator measuring the level of control of corruption in country *i*. It ranges from -2.5 to 2.5, where a higher value indicates better governance.	The World Governance Indicators database..
$Doil_{origin(t)}$	A binary variables equals one if exporter *i* is oil-abundant in terms of exporting petroleum, petroleum products and related materials in year *t*.	The UNCTAD Stat database.
$Dmineral_{origin(t)}$	A binary variables equals one if exporter *i* is mineral-abundant in terms of exporting ores and metals in year *t*.	The UNCTAD Stat database.

Appendix 2.6: China's ETCZs in Africa

Country	Name	Date of Establish-ment	Principal Sectors	Planned Investment (USD million)
Zambia	ZCCZ	2007	Copper products	800
Nigeria	Lekki FTZ	2007	Real estate development and logistics	330
Nigeria	Ogun State FTZ	2009	Electronic industry and furniture manufacturing	2500
Egypt	Suez SETZ	2009	Textile, petroleum, automotive and electrical appliance	460
Ethiopia	Dukem Industrial Park	2007	Textile, leather and construction equipment	849
Mauritius	Tianli ETCZ	2012	Logistics	220
Algeria	Mostaganem ETZ	2012	Automotive industry, construction materials and electronic industry	550
Botswana	Phalalane Industrial Park	2012		52

Source: we compile the information from the official websites for the FOCAC and Berthelemy (2011).

Appendix 2.7: Descriptive Statistics (Full Sample)

Full Sample	N	Max	Min	Median	Mean	Std. Dev.
Exports	105,840	3.6e+10	0.00	1310.82	4.1e+07	4.2e+08
$Ln(GDP_{origin})$	100,260	26.88	21.72	23.92	24.05	1.24
$Ln(GDP_{destination})$	102,735	30.21	18.71	23.99	24.09	2.26
$Ln(GDPpc_{origin})$	100,260	9.65	5.93	7.29	7.53	0.95
$Ln(GDPpc_{destination})$	102,651	11.25	4.94	8.62	8.56	1.28
$Ln(Dist)$	102,735	9.90	4.66	8.75	8.66	0.69
Common Border	105,840	1.00	0.00	0.00	0.02	0.15
$Landlocked_{origin}$	105,840	1.00	0.00	0.00	0.21	0.41
$Landlocked_{destination}$	105,840	1.00	0.00	0.00	0.21	0.41
Colonial Ties	105,840	1.00	0.00	0.00	0.01	0.08
Language	105,840	1.00	0.00	0.00	0.23	0.42
Legal	105,840	1.00	0.00	0.00	0.39	0.49
Currency Union	105,840	1.00	0.00	0.00	0.02	0.13
RTA	105,840	1.00	0.00	0.00	0.05	0.22

Appendix 2.8: Descriptive Statistics (Africa-China Sample)

Africa–China Sample	N	Max	Min	Median	Mean	Std. Dev.
Exports	224	2.1e+10	9348.02	1.1e+08	9.8e+08	2.8e+09
$Ln(GDP_{origin})$	215	26.88	22.22	24.23	24.34	1.22
$Ln(GDP_{China})$	224	29.84	29.11	29.50	29.48	0.25
$Ln(GDPpc_{origin})$	215	9.64	6.26	7.41	7.68	0.96
$Ln(GDPpc_{China})$	224	8.83	8.13	8.50	8.49	0.23
$Ln(Dist)$	224	9.47	8.93	9.30	9.25	0.15
$Landlocked_{origin}$	224	1.00	0.00	0.00	0.21	0.41
$Ln(FDI)_{origin(t-1)}$	196	21.74	12.35	17.26	17.16	1.77
$Duty\ Free_{origin}$	210	1.00	0.00	0.00	0.10	0.23
$Detcz_{origin}$	224	1.00	0.00	0.00	0.06	0.24
$Rule\ of\ Law_{origin}$	224	1.06	-1.82	-0.60	-0.62	0.60
$Control\ of\ Corruption_{origin}$	224	0.67	-1.48	-0.68	-0.63	0.48
$Doil_{origin}$	224	1.00	0.00	0.00	0.35	0.48
$Dmineral_{origin}$	224	1.00	0.00	0.00	0.33	0.47

Appendix 2.9: Correlation Matrix for the Full Sample

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
(1)Exports	1													
(2)Ln(GDP_{origin})	0.124	1												
(3)Ln($GDP_{destination}$)	0.169	0.016	1											
(4)Ln($GDP_{pcorigin}$)	0.089	0.516	0.011	1										
(5)Ln($GDP_{pcdestination}$)	0.099	0.018	0.465	0.010	1									
(6)Ln(Dist)	-0.037	-0.060	0.025	-0.027	0.204	1								
(7)Common Border	0.012	0.006	-0.045	-0.018	-0.146	-0.415	1							
(8)$Landlocked_{origin}$	-0.044	-0.266	0.001	-0.453	0.001	0.026	0.012	1						
(9)$Landlocked_{destination}$	-0.038	0.001	-0.130	0.002	-0.271	-0.159	0.062	-0.006	1					
(10)Colonial Ties	0.162	0.023	0.129	0.021	0.089	-0.044	0.035	-0.014	-0.042	1				
(11)Language	0.022	0.025	-0.193	0.017	-0.090	-0.139	0.147	0.063	-0.035	0.104	1			
(12)Legal	0.026	-0.016	-0.043	0.022	-0.080	-0.109	0.078	-0.031	-0.111	0.079	0.392	1		
(13)Currency Union	-0.003	-0.060	-0.076	0.011	-0.119	-0.258	0.197	-0.017	0.055	0.009	0.216	0.159	1	
(14)RTA	0.082	0.070	0.003	0.044	-0.084	-0.416	0.291	-0.033	0.026	0.045	0.079	0.047	0.176	1
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)

Appendix 2.10: Correlation Matrix for the Africa-China Sample

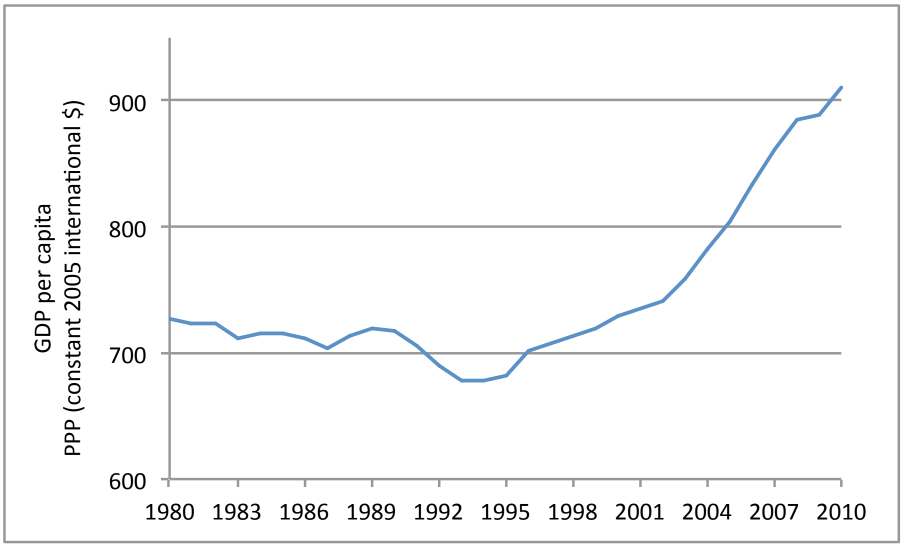
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
(1)Exports	1													
(2)Ln(GDP_{origin})	0.322	1												
(3)Ln(GDP_{China})	0.184	0.094	1											
(4)Ln($GDP_{pcorigin}$)	0.297	0.509	0.060	1										
(5)Ln($GDP_{pcChina}$)	0.184	0.094	1.000	0.060	1									
(6)Ln(Dist)	0.154	-0.374	0.000	-0.203	0.000	1								
(7)$Landlocked_{origin}$	-0.154	-0.253	0.000	-0.448	0.000	-0.052	1							
(8)Ln(FDI)$_{origin\ (t-1)}$	0.243	0.374	0.524	0.158	0.524	-0.056	-0.074	1						
(9)Duty $Free_{origin}$	-0.091	-0.180	0.247	-0.347	0.247	-0.100	0.387	0.117	1					
(10)$Detcz_{origin}$	-0.030	0.193	0.243	-0.085	0.244	-0.118	0.225	0.365	0.290	1				
(11)Rule of Law_{origin}	-0.156	0.014	-0.029	0.301	-0.029	-0.217	-0.024	-0.107	-0.007	-0.025	1			
(12)Control of $Corruption_{origin}$	-0.168	0.019	-0.052	0.253	-0.052	-0.060	-0.135	-0.011	-0.032	-0.032	0.887	1		
(13)$Doil_{origin}$	0.260	0.468	0.038	0.478	0.038	-0.183	-0.382	0.103	-0.292	0.044	-0.351	-0.441	1	
(14)$Dmineral_{origin}$	-0.015	-0.167	-0.003	0.008	-0.003	0.439	-0.048	0.177	0.023	-0.027	0.021	0.163	-0.300	1
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)

Chapter 3 Absence of structural change in African trade flows[29]

3.1 Introduction

Rapid economic growth is usually accompanied by structural change of the composition of export products. China, for example, has been growing rapidly since the 1980s, also after 2001 when it joined the World Trade Organization (WTO). Meanwhile, we observe dramatic changes in China's export structure from a primary agricultural products dominant structure to a manufactured goods based economy. Africa has also achieved relatively impressive economic performance recently. This is illustrated in Figure 3.1, which shows that GDP per capita has increased substantially since 1995. Does the economic growth performance of Africa lead to similar changes in the structure of African exports as in China? This study seeks to answer this question.

Figure 3.1: GDP per capita of Africa (1980-2010)



Source: ADI (African development indicator)

[29] This chapter is based on joint research with Charles van Marrewijk.

We provide an empirical framework to identify whether or not there is a structural break in African exports at a particular time during the last two decades. Based on all information from the entire distribution of revealed comparative advantages, we employ two indices to identify the evolution of African trade patterns. We also rule out pseudo structure breaks by analyzing individual countries. We find that African exports remained highly specialized on primary exports and there is no evidence of strong structural breaks in their exports over the last twenty years. In other words, the trade pattern of Africa remains concentrated on its traditional export sectors using low technology and limited human capital. All our findings suggest that the rapid economic growth of Africa is not associated with changes in the export structure of the countries involved.

We also analyze the causes of the absence of structural breaks of African exports. In principle, there are several forces promoting the mobility of international specialization, such as knowledge spillovers and technology transfer (Grossman and Helpman, 1991, Chapter 7). When lacking these drivers, however, the trade pattern is largely determined by initial conditions and reinforces itself over time (Lucas, 1988; Grossman and Helpman, 1991, Chapter 8). In the case of Africa, the absence of structural breaks in exports indicates that initial conditions largely determine the evolution of trade patterns. On the one hand, the abundance of endowment in land and natural resources contribute to the persistence of primary exports. On the other hand, the disadvantage of the labor force (low educational attainment), geographic and institutional factors impede the rise of manufacturing industries.

This paper contributes to the literature in two aspects. First, we investigate recent changes in African trade structures by analyzing (the absence of) structural breaks in African export patterns. We will hardly find evidence for such structural breaks compared to similar analyses done for other countries. Second, we discuss some possible causes of the absence of structural change. The remainder of this paper is organized as follows. Section 3.2 reviews the related literature on revealed comparative advantage and African trade flows. Section 3.3 sets out the methodology used to identify structural changes. Section 3.4 discusses the data sources. Section 3.5 presents the results on identifying structural breaks in African comparative advantage for export flows. Section 3.6 discusses some plausible reasons for the African persistence in comparative advantage. Section 3.7 provides concluding remarks.

3.2 Related studies

Since the concept of comparative advantage is defined using unobserved relative prices under autarky researchers have measured comparative advantage in a more indirect way. Most prominent, and most often used in this respect, is the concept of revealed comparative advantage (henceforth, RCA), also known as the Balassa index (Balassa, 1965).

Based on summary statistics regarding the distribution of the Balassa index, various empirical studies identify the evolution of trade patterns by either comparing the standard deviation of Balassa indices of each sector over time or putting Balassa indices of the first and final year into an OLS regression. For instance, by calculating Balassa indices of 73 manufacturing sectors, Balassa (1977) analyzes the trends of trade specialization in industrialized countries from 1953 to 1971. Using revealed comparative advantage as the dependent variable, Balassa (1979) estimates whether a country has comparative advantage in capital-intensive products in 184 manufactured goods in 36 developed and developing countries. He continues the study in the trend of a country's comparative advantage along with the accumulation of physical and human capital endowments. Dalum, Laursen and Villumsen (1998) apply a modified Balassa index, the revealed symmetric comparative advantage (henceforth, RSCA), into an OLS regression model to analyze whether there is a high degree of stability of export specialization in OECD countries from 1965 to 1992. The regression of the RSCA of the final year on that of the initial year implies that the country tends to become more specialized in sectors of which the estimated coefficient is greater than one, and vice versa. They find that OECD countries show a declining trend in export specialization, except for Greece, Italy, Japan and the USA.

More recent literature focuses on the entire distribution of Balassa indices rather than summary statistics. Using Markov transition probability matrices and mobility indices, Proudman and Redding (2000) evaluate the persistence or mobility of the distribution of Balassa indices over time, which reflects the evolution of trade patterns. They analyze differences in the dynamics of international trade patterns for Japan, France, Germany, the UK, and the USA. There is no increasing trend in international specialization in most countries except Japan. By contrast, Japan has a high degree of persistence in the changes of the revealed comparative, suggesting that there is an increasing trend of international specialization. Expanding Proudman and Redding's analysis by augmenting the sample of six largest industrialized countries and eight fast growing Asian economies, Brasili, Epifani and Helg (2000) apply stochastic kernels to evaluate the dynamics of trade patterns, on top of the Markov transition probability matrix. Hinloopen and van Marrewijk (2001) systematically analyze the empirical distribution of the Balassa index in 12 European countries. Based on the calculations of Markov transition probability matrix and four types of mobility indices in monthly and annual data, respectively, they find that Germany has the most persistent pattern while Greece has the most mobile pattern of comparative advantages in 1996.

In addition to the previous method, Hinloopen and van Marrewijk (2005) develop the Harmonic Mass index (HM index), which estimates the dynamics of the whole distribution of Balassa indices over time, to capture structural changes in trade patterns. This index was applied in some empirical research. Hinloopen and van Marrewijk (2004) capture two substantial structural changes (the 1980-1985 transition and the 1985-1990 transition) of China's export flows during 1970-1997 by using the HM index. Brakman, Inklaar and van Marrewijk (2013) use the HM index in the export flows of the OECD countries. They find a substantial structural change of the OECD countries in the middle of the 1980s and attribute the structural break to the large fluctuation in the real effective exchange rate of the dollar in the 1980s. The methodology was further developed to the Harmonic Weighted Mass (HWM) index by Hinloopen, Wagenvoort and van Marrewijk (2012), which is the method we use in this paper.

Several papers use revealed comparative advantage to analyze export patterns in Africa (Ariovich, 1979; Ng and Yeats, 2001; Edwards and Schoer, 2002). The results of static and dynamic analyses suggest that the comparative advantage concentrates on the traditional exports sectors, such as agricultural products and natural resource intensive product. The shift of comparative advantage from traditional products to manufactured goods is difficult due to the rich endowment in natural resources in Africa. In addition to the abundant natural resources, Wood and Berge (1997) and Wood and Mayer (2001) argue that the low level of education, poor infrastructure and bad policies in Africa are the main explanations for the primary products-dominant in the export structure of Africa.

3.3 Methodology

Traditional methods identify the export structure of a country by comparing the share of each sector in total exports. Though this method gives a general idea about the comparative advantages of different sectors, it is not an accurate proxy of comparative advantage. Besides, it is not applicable in a cross-country comparison, since it is not necessary for a sector with a high share in export in country A to have comparative advantages over the same sector in country B with an even higher share. In this paper, relying on the concept of revealed comparative advantage, we identify the African export structure by calculating the Balassa index, which is success in measuring the comparative advantage of a particular sector in a country and frequently used in the previous literature. Moreover, based on the entire distribution of the Balassa index, we further calculate the Harmonic Weighted Mass (henceforth, HWM) index to capture the dynamic of the trade patterns. To supplement the statistical analysis of the HWM index and further confirm the findings, we use an alternative method associated with the Markov transition probability matrix and mobility indices to analyze the dynamics of revealed comparative advantage in Africa. Because their calculation also depends on the entire distribution of the Balassa index and often used in the literature. In this section, we discuss the indices used.

3.3.1 Balassa index (BI)

Comparative advantage is an important concept in international trade theory. However, it is not easy to identify the true theoretical comparative advantage due to unobservable variables in empirical work (De Benedictis and Tamberi, 2001). It is thus useful to introduce the concept of revealed comparative advantage to empirically analyze specialization patterns in a cross-country setting. Balassa (1965) develops the mostly popular measure of revealed comparative advantage, the Balassa index, for country *c* in sector *s*, defined as

$$BI_{cs} = \frac{Export_{cs}/Total\ export_{c}}{Export_{refs}/Total\ export_{ref}} \tag{1}$$

where *ref* denotes a group of reference countries. The revealed comparative advantage of sector *s* is given by the ratio of its share of exports in country *c* to its share of exports in the reference countries.[30] If BI_{cs}>1, namely country c displays a sectoral share greater than the same sectoral share in the reference countries, we conclude that the sector s in country c has revealed comparative advantage over the reference countries, and vice versa. Its congruence with the theoretical foundation was developed later (Hillman, 1980). Based upon exports as the only information variable, the Balassa index has a direct interpretation, making it a convenient tool to calculate and apply in empirical work.

The Balassa index is not symmetric and its mean value is not fixed. Hence, some researchers propose to normalize the original Balassa index in various ways (Laursen, 1998; Proudman and Redding, 2000). For instance, Laursen (1998) develops the normalization as $BI_L = \frac{BI-1}{BI+1}$ to deal with the skewness of the Balassa index, where BI_L ranges from -1 to +1 and is therefore symmetric. When $-1 < BI_L < 0$, the sector s has a comparative disadvantage in country c. By contrast, country c has a comparative advantage in sector s if $0 < BI_L < 1$. De Benedictis and Tamberi (2001), however, claim that the normalization is inappropriate as it induces a bias arising from extreme values of the distribution. Weighting the original index by the number of sectors n, Proudman and Redding (2000) normalize the Balassa index as $BI_P = nBI$. Therefore the mean of the normalized Balassa indices of all sectors is equal to one. The drawbacks of this normalization are also discussed in De Benedictis and Tamberi (2001). This normalization leads to serious errors of interpretation, especially when the Balassa index distribution is concentrated around the mean. Since both normalizations have their drawbacks, we use the original Balassa index given in Equation (1) to define revealed comparative advantage of African exports.

3.3.2 Harmonic Weighted Mass index (HWM)

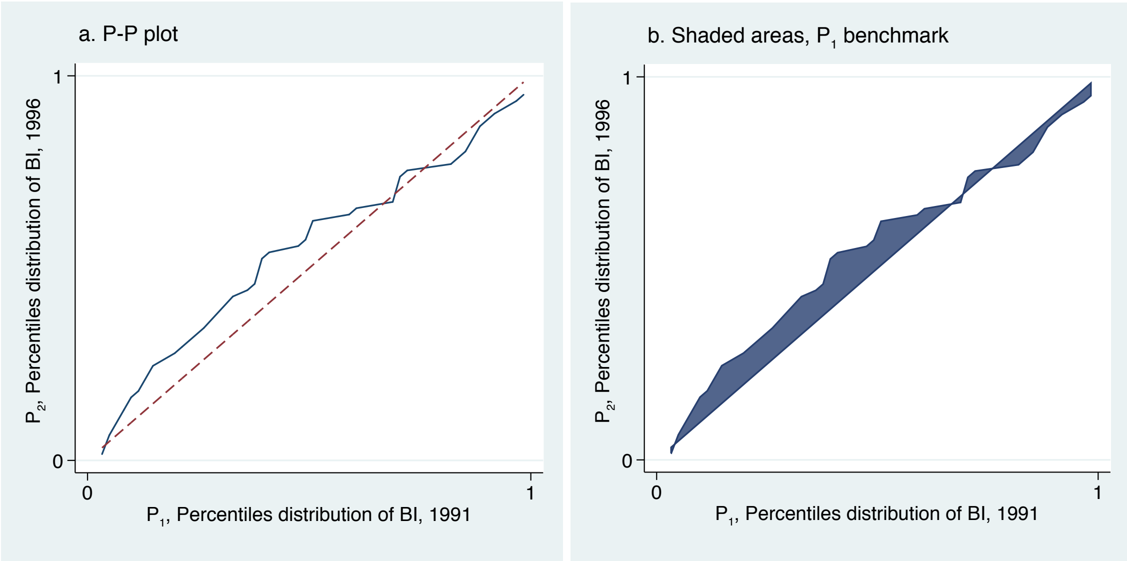
Since the Balassa index is a static statistic, which cannot give full information about the dynamic evolution of the entire Balassa indices over time. A number of alternative statistics are developed in the literature, such as stochastic kernel estimates and Markov transition matrices (see the next subsection), in spite of practical drawbacks. The estimation of stochastic kernels is complicated and requires a large number of data. Most importantly, its results are of limited practical use (Hinloopen and van Marrewijk, 2004).

[30] Usually the world or some special regions are used as the reference group.

When calculating Markov transition matrices, the Balassa index values have to be divided into distinct discrete cells, increasing calculation burden. In this paper, we use the HWM index to identify structural breaks in African trade patterns. The key advantage of the HWM index is that a series of Balassa index *distributions* can be compared over time using a single statistic. More importantly, the 10 percent critical value of the HWM is approximately 0.5. This value is virtually invariant with the number of observations (there is limited variation for low numbers of observations), so this cutoff makes it easy to determine significance.

We first introduce the probability-probability plots (P-P plots), the basis of the calculation of the HWM index. A P-P plot is used to assess whether two data sets come from an identical distribution by plotting the two cumulative distribution functions against each other. We select a set of Balassa indices of all sectors in a particular country for a particular year as one data set and compare it to a set of Balassa indices of all sectors in the same country 5 years later (or 1, 2, 3, or 4 years later). The P-P plot shows how the distributions of Balassa indices of all sectors in a country change over time. Taking Angola for example, Figure 3.2a shows the P-P plot of the Balassa index in 1991 and 1996, respectively. We take the Balassa indices of all 2-digit commodities in Angola in 1991 and 1996 as two data sets, respectively. Let $F(x)$ be the joint cumulative distribution functions (CDF) of the two data sets under the assumption that the two datasets are from the same distribution. The empirical cumulative distribution functions $F_1(x)$ and $F_2(x)$ correspond to the data sets in 1991 and 1996, respectively. P_1 and P_2 denote the probabilities that a random variable takes on a value smaller than or equal to x in the functions $F_1(x)$ and $F_2(x)$. We can compare the two cumulative distribution functions on a compact scale by defining $P_2 = F_2(F_1^{-1}(P_1))$ and then make a P-P plot based on P_1 and P_2 (see Figure 3.2a). If $P_1 = P_2$, then the two cumulative distribution functions are identical and the plot coincides with the 45° dashed line. As for the case of Angola, the distributions of the Balassa indices of all 2-digit commodities are apparently different between 1991 and 1996. The HWM index provides a method to determine whether this difference is statistically significant.

Figure 3.2: P-P plot of Balassa index (Angola, 1991-1996)



Since we are interested in the changes in trade patterns over time, we are particularly concerned with the magnitude of the changes. By calculating the size of the shaded areas in Figure 3.2b, we identify the extent of these changes. That is the basic idea of the HWM index. Since the function of the 45° dashed line is $y = P_1$, the shaded area between P-P plot and $45°$ dash line is calculated by the integral $\int_0^1 |P_2 - P_1|\, dP_1$. The HWM index weighs the size of the shaded areas by the number of observations of two data sets and is defined as

$$HWM = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \int_0^1 |P_2 - P_1|\, dP_1 \qquad (2)$$

Where n_1 and n_2 are the numbers of observations of each data set, respectively.

It is quite straightforward to illustrate the HWM index. If the HWM index is greater than the critical value 0.5, we conclude that the two data sets are derived from *different* underlying distributions. In that case, we will say a *structural change* has occurred. If the value is smaller than 0.5 we cannot draw this conclusion. The result of our Angola example is a HWM index of 0.258 only. Hence, we find no evidence that the trade pattern in Angola changed significantly from 1991 to 1996. We use this procedure using the Balassa indices of 2-digit export commodities in African countries from 1991 to 2009 to test whether their Balassa indices are drawn from the same distribution in 1-5 year differences.

3.3.3 Markov transition probability matrix and mobility index

In this subsection, we discuss the *Markov transition probability matrix* and *mobility indices*. Before showing these methods, let us introduce some basic concepts. A *Markov process* is a random process in which only the current state of the process influences where it goes next. In other words, given the present, the future is independent of the past. The *Markov process* is known as a *Markov chain* if both the time space and the state space are discrete. A *Markov Chain* is a mathematical system that goes through transitions from one state to another. It can be summarized by a transition matrix, which is widely used in social science applications, such as the convergence of the income level and the dynamics of the export structure (see Quah, 1993; Proudman and Redding, 2000; Hinloopen and van Marrewijk, 2001). In this paper, we take the discrete time data of the Balassa indices of twenty African countries as *Markov chains* to identify their persistence and mobility.

A. Markov transition probability matrix

We consider a *Markov chain* $\{BI_{cst}, t = 0, 1, 2, \dots\}$ as a nonnegative set, where BI_{cst} is the Balassa index of a particular sector s in a particular African country c at time t. Unless otherwise mentioned, we use BI_t as a replacement of BI_{cst}. All the elements of the *Markov chain* are grouped into several states. If the element $BI_t \in i$, then the process is said to be in state i at time t. We suppose that the process makes a transition from state i to the next state j with a fixed probability P_{ij}. In particular

$$P\{BI_{t+1} = j | BI_t = i, BI_{t-1} = i_{t-1}, \dots, BI_1 = i_1, BI_0 = i_0\} = P_{ij} \tag{3}$$

for all states $i_0, i_1, \dots, i_{t-1}, i, j$ and all $t \geq 0$. Equation (3) states that, given the past states $BI_0, BI_1, \dots, BI_{t-1}$ and the present states BI_t, the conditional distribution of any future state BI_{t+1} is independent of the past states and depends only on the present state. Since the Probability P_{ij} is nonnegative and the process must move into some state from the current state, we have that

$$P_{ij} \geq 0, \quad i, j = 1, 2, \dots, m; \quad \sum_{j=1}^{m} P_{ij} = 1, \quad i = 1, 2, \dots, m \tag{4}$$

Let P denote the $m \times m$ matrix of one-step transition probabilities P_{ij}, that is:

$$P = \begin{bmatrix} P(1,1) & P(1,2) & \cdots & P(1,m) \\ P(2,1) & P(2,2) & & P(2,m) \\ \vdots & & \ddots & \vdots \\ P(m,1) & P(m,2) & \cdots & P(m,m) \end{bmatrix}$$

The entries on the diagonal of the matrix are the probabilities of remaining in the same state from present time t to the next period $t+1$. In this sense, they identify the *persistence* of the Markov chain. The entries on the off-diagonal of the matrix indicate the transition probabilities of moving from the present state BI_t to the future state BI_{t+1}. For example, $P(1,2)$ in matrix *P* indicates the probability of the Balassa index moving from state 1 at time *t* to state 2 at time *t+1*. In this paper, we categorize all Balassa indices into four states (see details in 3.5), so $m = 4$ and *P* is a 4×4 matrix.

B. Limiting probability distributions

Matrix *P* is the one-step transition probability matrix estimated from the Markov Chain of a series of Balassa indices. Such a one-step transition is labeled a 1-year transition. Additionally, a long-run transition of this process is worth noting. We denote P_{ij}^t as the *t*-step transition probability and P^t is the corresponding matrix. If the one-step transition matrix *P* is irreducible, recurrent and aperiodic (see the definitions in Ross, 2000: P169, P172 and P178, respectively), P_{ij}^t will converge to some value (π_j) as *t* approaches infinity, which is the same for all *i*. In other words, π_j is the probability of reaching state *j* after a large number of transitions regardless of the state status of the initial process. The limiting probability distribution (π_j) is called the *ergodic distribution*. The theoretical ergodic distribution can be characterized by:

$$\pi_j = \lim_{t\to\infty} P_{ij}^t = \sum_{i=1}^{4} \pi_i P_{ij}, \; j = 1, 2, 3, 4 \qquad \sum_{j=1}^{4} \pi_j = 1 \tag{5}$$

The above two equations (5) uniquely determine the stationary probability distribution π_j as $t \to \infty$. Alternatively, we can simply iterate the one-step matrix *P* until each column (*j*) converges to the stationary status (π_j). In this paper, we iterate the one-step transition probability matrix *P* for 10,000 times to calculate the theoretical ergodic distributions. Unfortunately, not every country in our sample has an ergodic distribution since the transition probability matrices of some countries do not satisfy the irreducible, recurrent and aperiodic conditions. This holds for Algeria, Libya, Mali, and Mauritius (see Appendix 3.5).

We also calculate the share of Balassa indices in each state during the entire sample period, which are labeled *empirical* ergodic distributions. By comparing the results of the *theoretical* and *empirical* ergodic distributions, we get an indication if the transition probability matrices accurately capture the data-generating process underlying the Balassa index distributions.

C. Mobility index

By reducing the information captured by transition probability matrix *P* to a single statistic, a *mobility index* characterizes the degree of mobility of the Balassa index process. Various mobility indices have been developed in the literature. Since there is no significance test for these mobility indices, we have to compare our results to some benchmarks. In this paper, we follow the calculation of four mobility indices in Hinloopen and van Marrewijk (2001) and Proudman and Redding (2000) and adopt their results as benchmarks.

Table 3.1: Four Mobility Indices

	Literature	Mobility Index	Description
1	Shorrocks (1978)	$M_1 = \frac{m - tr(P)}{m-1}$	• m is the number of states (m=4); • $tr(P)$ is the trace of the one-step transition probability matrix *P*.
2	Bartholomew (1973)	$M_2 = \sum_j \pi_j \sum_i P_{ji} \lvert j - i \rvert$	• π_j is the ergodic distribution of transition matrix *P*; • *i, j* are the state; • P_{ji} is the probability of one-step transition form state j to state *i*.
3	Sommers and Conlisk (1979)	$M_3 = 1 - \lvert \lambda_2 \rvert$	• λ_2 is the second largest eigenvalue of matrix *P*.
4	Shorrocks (1978)	$M_4 = 1 - \det(P)$	• det (*p*) is the determinant of matrix *P*.

Table 3.1 shows four methods of calculating mobility indices. First, the diagonal elements of M_1 are the probability of staying in the same state. One minus these elements is thus a measure of mobility. Second, M_2 uses the proportions π_j of the ergodic distribute as weights to calculate the degree of mobility of the off-diagonal elements. Third, let $\lambda_1, \dots, \lambda_n$ be the eigenvalues of the one-step transition matrix P, ordered by $\lambda_1 = 1 > |\lambda_2| \geq \cdots \geq |\lambda_n|$. The t-step transition probability matrix P^t converges to its equilibrium value (π_j) as rapidly as $|\lambda_2|^t, \dots, |\lambda_n|^t$ converge to zeros (Sommers and Conlisk, 1979). The second largest eigenvalue, $|\lambda_2|$, is commonly used as a measure of convergence speed or a measure of immobility, because $|\lambda_2|^t$ converges as slow or slower than the other eigenvalues. Thus M_3 uses 1 minus the absolute value of the second largest eigenvalue as an indication of mobility. Fourth, and finally, the product of the eigenvalues is equal to the determinant of the transition matrix P, which explains M_4.

The advantage of calculating four types of mobility indices is that, by adopting some properties of the transition matrix, we are able to evaluate mobility through the entire distribution of Balassa indices. When these indices are positively correlated, the four mobility indices properly capture the degree of mobility. Therefore, we can simply compare our results with the benchmark, identifying whether there is a reasonable structural change in African exports.

3.4 Sample and data sources

By 2010, there are 57 countries in Africa (United Nations, 2011). Because not every country plays an important role in African exports, we restrict our sample according to the following criteria. First, a country is included if it has a population of more than one million people in 2010. Second, we select countries with available annual data for more than 10 years, which are well distributed during the period from 1990 to 2010. Third, the total value of exports from the selected countries should rank top twenty in Africa either in

1990, 2000 or 2010. In the end, our sample consists of twenty African countries (Appendix 3.1).[31] In addition, 181 countries are used as our reference countries (see Appendix 3.2). We collect annual export data from 1990 to 2010 using UN Comtrade.[32] The export data includes 69 sectors at the 2-digit level of the classification SITC Rev.2, which contains 99.998% of all African exports in our sample. Moreover, to cope with the volatile data in individual years, we use 3-year central moving average to the annual export data. Consequently, the time dimension of our sample reduces to 19 years.

3.5 Results of identifying structural change

3.5.1 Results with Balassa index

We calculate the Balassa index for 20 African countries in the period 1991-2009. In order to identify changes in the comparative advantage of major exports in each African country, we list the 2-digit sectors with the greatest Balassa index in 1995 and 2005, respectively (Table 3.2). Thirteen countries showed no shift in the sectors with the largest Balassa index from 1995 to 2005. In particular, Ethiopia, Kenya, Madagascar and Uganda always had the largest comparative advantage in sector SITC 07 (Coffee, tea, cocoa, spices, and manufactures thereof). Similarly for some natural-resource products in several countries, such as Morocco, South Africa, Angola, Libya, Algeria and Zambia. Almost one-third of the countries (seven in total), however, showed a change in the 2-digit sectors with the highest Balassa index. Cote d'Ivoire, for example, shifted from an agricultural product (SITC 07) to a technology-intensive product (SITC 95). The other countries had their largest comparative advantage always either in agricultural or natural-resource sectors in these two years.

[31] There are two exceptions, Angola and Libya, which do not meet the above requirements but are included. There is not enough export data for Angola and Libya in the UN Comtrade if taking them as reporter countries. But these two countries are important in African exports, as they are the largest exporters to China and the USA in 2010, respectively. So we reverse the reporter and partner countries in order to get their exports data, taking top 50 importers as the reporter countries, which accounted for 92.73% of the total imports in the world in 2010, and on the other hand taking Angola and Libya as partner countries.

[32] http://comtrade.un.org/

Table 3.2: The highest Balassa index in selected years, 2-digit

Country	Year	SITC	Commodity	BI
Ethiopia*	1995	07	Coffee, tea, cocoa, spices, and manufactures thereof	105.6
	2005			96.7
Kenya	1995	07	Coffee, tea, cocoa, spices, and manufactures thereof	57.9
	2005			51.0
Madagascar	1995	07	Coffee, tea, cocoa, spices, and manufactures thereof	59.0
	2005			28.6
Uganda	1995	07	Coffee, tea, cocoa, spices, and manufactures thereof	103.8
	2005			63.5
Malawi	1995	12	Tobacco and tobacco manufactures	121.1
	2005			220.6
Gabon*	1995	24	Cork and wood	17.4
	2005			15.6
Morocco	1995	27	Crude fertilizer and crude minerals	27.2
	2005			25.1
South Africa	1995	32	Coal, coke and briquettes	12.9
	2005			12.7
Angola	1995	33	Petroleum, petroleum products and related materials	18.8
	2005			9.6
Libya	1995	33	Petroleum, petroleum products and related materials	17.8
	2005			9.4
Algeria	1995	34	Gas, natural and manufactured	44.0
	2005			21.5
Tunisia	1995	56	Fertilizers, manufactured	17.7
	2005			18.9
Zambia*	1995	68	Non-ferrous metals	41.7
	2005			28.3
Cote d'Ivoire*	1995	07	Coffee, tea, cocoa, spices, and manufactures thereof	69.2
	2005	95	Armoured fighting vehicles, war firearms, ammunition, parts, nes	97.6
Egypt	1995	94	Animals, live, nes, (including zoo animals, pets, insects, etc)	14.3
	2005	26	Textile fibres (not wool tops) and their wastes (not in yarn)	11.0
Mali*	1995	26	Textile fibres (not wool tops) and their wastes (not in yarn)	145.8
	2005	97	Gold, non-monetary (excluding gold ores and concentrates)	149.3
Mauritius	1995	06	Sugar, sugar preparations and honey	69.1
	2005	94	Animals, live, nes, (including zoo animals, pets, insects, etc)	215.2
Mozambique	1995	03	Fish, crustacean and molluscs, and preparations thereof	45.2
	2005	35	Electric current	31.5
Senegal	1995	27	Crude fertilizer and crude minerals	32.7
	2005	03	Fish, crustacean and molluscs, and preparations thereof	28.7
Togo*	1995	27	Crude fertilizer and crude minerals	88.7
	2005	26	Textile fibres (not wool tops) and their wastes (not in yarn)	49.5

Note: * indicates there is no data for 1995 or 2005, the data of the closest year is listed in the table

We are not primarily interested in the changes in the sector with the largest Balassa index in a country, but mostly with the change in comparative advantages between industries within the country. We classify the 2-digit SITC commodities into five factor-abundance type categories based on the *Empirical Trade Analysis Center*.[33] We made some adjustments in the classification for two reasons. First, our data is collected at the 2-digit instead of the 3-digit level. Second, the majority of African exports concentrate in *primary products*. Thus we divided the first group (A. primary products) of the Empirical Trade Analysis Center, into two parts, namely *Agricultural primary products* and *Natural resource & natural-resource intensive products*. Specifically, except for the sectors SITC27, SITC28, and SITC3 in the original classification the remaining products are grouped into the first type: Agricultural primary products. In addition, the sectors SITC 27, SITC 28, SITC 3 and SITC 971 and all sectors in the second group (B. Natural-resource intensive products) in the original classification are combined in the second type in this paper: Natural resource & Natural-resource intensive products. The complete classification is listed in Appendix 3.3 and Appendix 3.4.[34]

- A: Agricultural primary products (23 sectors)
- B: Natural resource & natural-resource intensive products (11 sectors)
- C: Unskilled-labor intensive products (6 sectors)
- D: Technology intensive products (16 sectors)
- E: Human-capital intensive products (10 sectors)

For each of these groups, we calculate the share of sectors that have a Balassa index greater than unity in the sectors. The idea is that changes in these shares are expected to reflect changes in trade patterns. Table 3.3 shows the information for two selected years, 1995 and 2005. The shaded numbers contain the largest shares in each year, implying that the comparative advantage of the trade pattern concentrates on this group. As expected, most African countries in our sample had a comparative advantage in Agricultural Primary Product than other groups in both years. As for Algeria, Angola, Gabon and Libya, the small share of the comparative advantageous sectors in the Natural resource and natural-resource intensive products indicates that they have some strong sectors in their exports, in

[33] Website http://www2.econ.uu.nl/users/marrewijk/eta/index.htm

[34] There are three sectors (SITC91, 93 and 96) grouped into Non-classified products, which is not listed here.

line with their high dependency on oil and natural gas exports. Comparing the distribution of the sectors with BI>1 among different groups, we find that there was no shift of comparative advantage among groups from 1995 to 2005 in our sample countries.

Table 3.3: Share of sectors with BI>1 in total sectors, percentage (%)

Country	Agricultural Products		Natural-resource		Unskilled-labor int.		Technology int.		Human-capital int.		Total Share	
	1995	2005	1995	2005	1995	2005	1995	2005	1995	2005	1995	2005
Cote d'Ivoire*	17.4	17.4	4.3	2.9	0.0	0.0	1.4	2.9	1.4	2.9	24.6	26.1
Egypt	8.7	10.1	5.8	8.7	4.3	2.9	2.9	2.9	2.9	1.4	24.6	26.1
Ethiopia*	15.9	17.4	1.4	2.9	0.0	0.0	0.0	0.0	0.0	0.0	17.4	20.3
Kenya	20.3	15.9	5.8	5.8	0.0	2.9	0.0	2.9	2.9	2.9	29.0	30.4
Madagascar	14.5	14.5	2.9	4.3	1.4	2.9	0.0	0.0	1.4	2.9	20.3	24.6
Malawi	13.0	11.6	0.0	1.4	0.0	1.4	0.0	1.4	0.0	0.0	13.0	15.9
Mali*	5.8	4.3	2.9	1.4	0.0	0.0	0.0	0.0	0.0	0.0	8.7	5.8
Mauritius	5.8	5.8	0.0	1.4	2.9	2.9	2.9	1.4	0.0	2.9	11.6	14.5
Morocco	8.7	11.6	5.8	4.3	5.8	5.8	4.3	4.3	0.0	0.0	24.6	26.1
Mozambique	11.6	13.0	4.3	4.3	0.0	0.0	0.0	0.0	1.4	0.0	17.4	17.4
Senegal	11.6	18.8	2.9	4.3	0.0	0.0	4.3	5.8	1.4	1.4	20.3	30.4
South Africa	11.6	14.5	10.1	8.7	1.4	1.4	4.3	5.8	1.4	2.9	29.0	33.3
Togo*	13.0	18.8	2.9	2.9	0.0	0.0	0.0	0.0	0.0	5.8	15.9	27.5
Tunisia	5.8	7.2	4.3	4.3	4.3	5.8	4.3	4.3	0.0	0.0	18.8	21.7
Uganda	14.5	18.8	1.4	4.3	0.0	0.0	1.4	0.0	0.0	2.9	17.4	26.1
Zambia*	7.2	14.5	4.3	5.8	1.4	1.4	2.9	2.9	0.0	1.4	15.9	26.1
Algeria	0.0	0.0	4.3	2.9	0.0	0.0	0.0	0.0	0.0	0.0	4.3	2.9
Angola	0.0	0.0	1.4	2.9	0.0	0.0	0.0	0.0	0.0	0.0	1.4	2.9
Gabon*	2.9	4.3	4.3	4.3	0.0	0.0	0.0	0.0	0.0	0.0	7.2	8.7
Libya	1.4	0.0	2.9	2.9	0.0	0.0	2.9	1.4	0.0	0.0	7.2	4.3

Note: * indicates there is no data for 1995 or 2005, the data of the closest year is listed in the table
The shaded cells contains the largest shares in each year; total share indicates the total share of sectors with a revealed comparative advantage (Balassa index > 1).

The *total share* indicates the overall share of sectors with a revealed comparative advantage (Balassa index > 1) out of a total of 69 sectors (see Table 3.3). Except Senegal, Uganda and Zambia, there is virtually no change in this share. It is worth noting that the average share of sectors with a revealed comparative advantage in Africa is 16 percent in 1995 and less than 20 percent in 2005 (not shown in the table). This is much lower than that for other economies, such as 33 percent in EU-12 from 1992 to 1996 and 40 percent in China from 1970 to 1997.[35] This is indicative of the fact that African comparative advantage is more concentrates in a small number of sectors. The next subsection analyzes

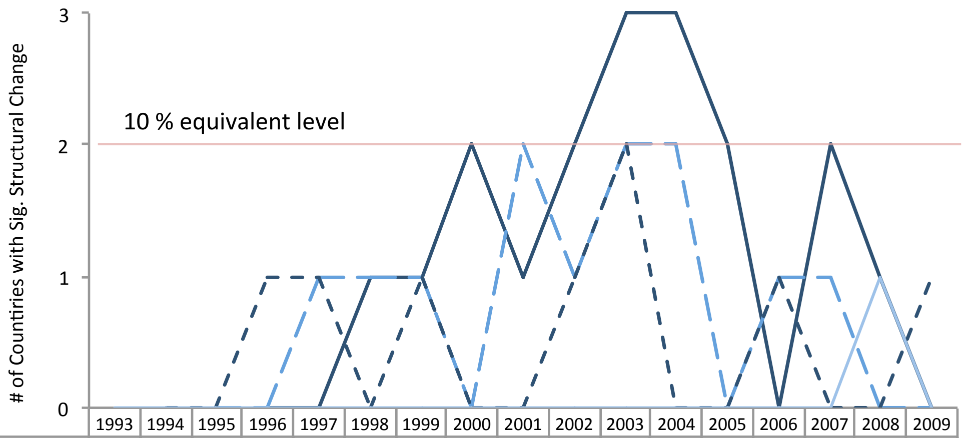
[35] See Hinloopen and van Marrewijk (2001 & 2004).

the whole distribution of the Balassa indices within each African country from 1991 to 2009.

3.5.2 Results for the HWM index

By comparing the distribution of the Balassa index at two points in time using the Harmonic Weighted Mass (HWM) index we can identify structural changes as explained in section 3.3. We calculate the HWM index for each country in our African sample by starting in 1991 and comparing the distribution of the Balassa index with that in 1992, 1993, 1994, 1995, and 1996 (five subsequent years). We then move on to 1992 and repeat the exercise. We keep doing this until we finally compare 2004 with 2005, 2006, 2007, 2008 and 2009.

Figure 3.3: The number of African countries with significant structural change

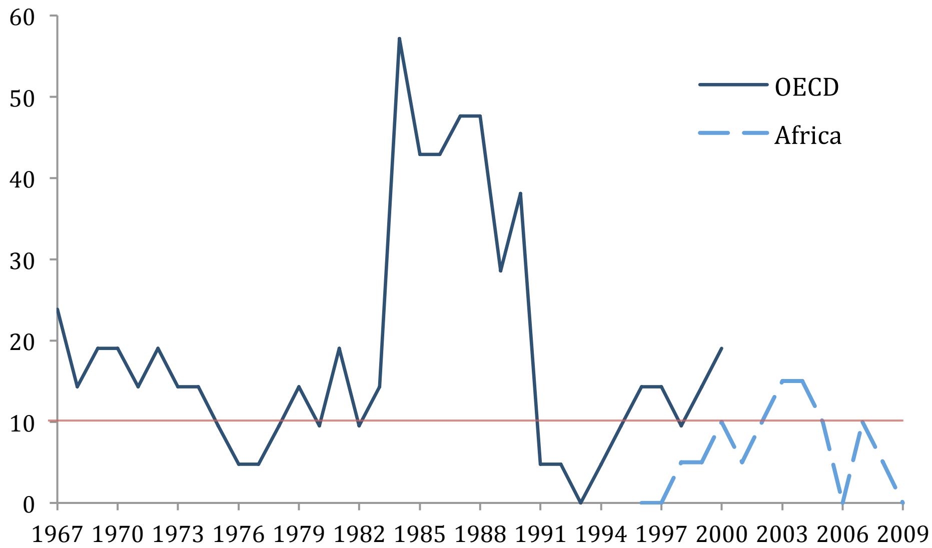


	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
5-year				0	0	1	1	2	1	2	3	3	2	0	2	1	0
4-year			0	0	1	1	1	0	2	1	2	2	0	1	1	0	0
3-year		0	0	1	1	0	1	0	0	1	2	0	0	1	0	0	1
2-year	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0

Note: The reported difference is backward in time; 3-year moving average; Significance at 10% level.

Figure 3.3 presents the number of African countries with significant HWM indices for 1-5 year differences in the sample period (using the 10 percent significance level as a cut off). The number of structural changes is only ‘high’ if it is above the number that we should normally expect. With a 10 percent significance level and 20 countries that benchmark number is, of course, 2 countries, as indicated in the figure. The actual number is always below this threshold, except for 5-year differences in 2003 and 2004. Note that the average number over the whole period is also below the threshold for 5-year differences. In short, we find only a few countries with structural breaks using 2-5 year differences during the entire sample period and no countries with a structural break using 1-year differences. Overall, the number of structural breaks does not exceed the expected threshold level during the time span.

Figure 3.4: The share of the countries with structural change in each sample (%)[36]



Note: The reported difference is backward in time; 5-year difference; significance at 10% level; 5-year moving average in OECD, 3-year moving average in Africa.

How do our results for Africa compare to other studies? Brakman, Inklaar and van Marrewijk (2013), use a similar method to analyze structural change of comparative advantage for OECD countries. Their results for 5-year differences are summarized and compared to our case in Figure 3.4. Clearly, structural change is much more frequent in OECD countries. The overall frequency is clearly above the 10 percent threshold (see also below), which is also the case in most individual years. There is a peak in structural breaks in the OECD countries in the 1980s.[37] In contrast, structural change in Africa is nearly absent. Table 3.4 shows the share of countries with significant changes for various years of difference for both the OECD and African countries during the entire sample period. Structural change is higher for longer time differences in both cases, as expected. In all cases the share of structural change is substantially higher in the OECD sample than in

[36] The two studies include 21 OECD countries and 20 African countries, respectively.

[37] The authors analyze several possible causes of this peak, see Brakman, Inklaar, and van Marrewijk (2013).

Africa. More specifically, this share exceeds the 10 percent threshold in the OECD countries for 3-5 year differences. This is never the case in Africa in the past two decades.

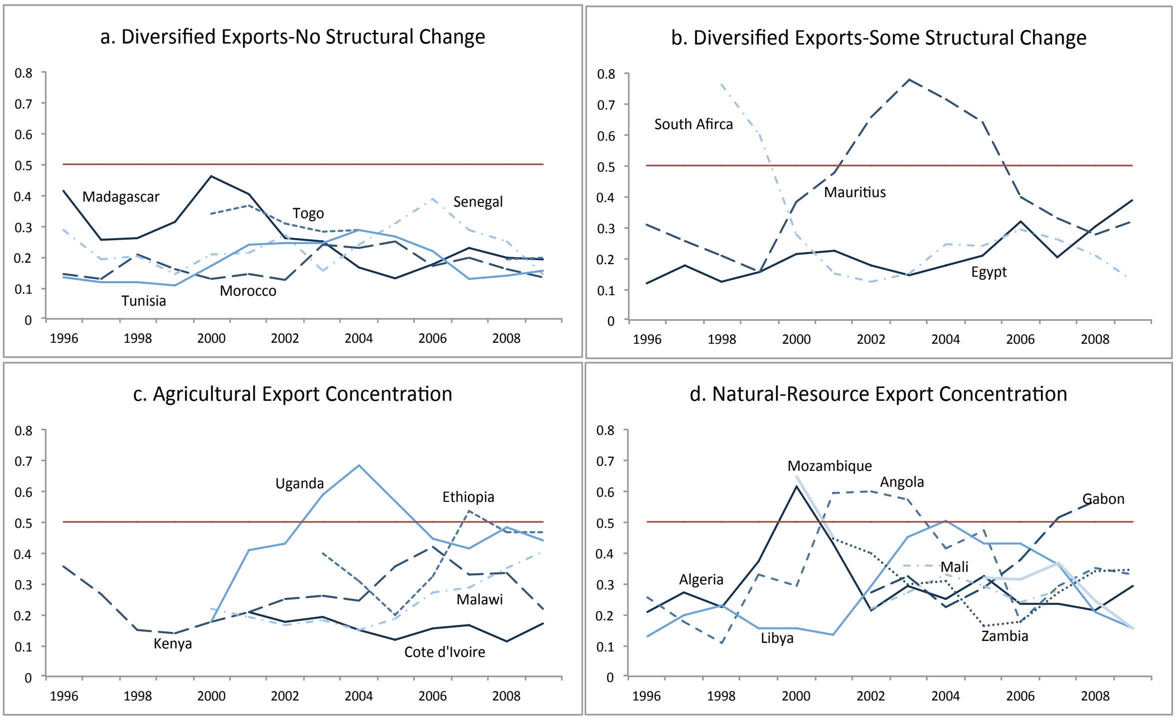
Table 3.4:The share of countries with structural changes in each# of years difference (%)[38]

Regions	5-year	4-year	3-year	2-year	1-year
OECD	18.63	14.29	10.05	5.53	1.75
Africa	6.43	4.00	2.50	0.29	0.00

Figure 3.5 shows the results of the HWM index calculations at the country level. For tractability, we restrict attention to the (largest) 5-year differences and display 3-year moving averages. Since a country experiences structural change if the HWM index is greater than 0.5 we first note that there is a large number of 11 African countries without any structural change in the 5-year difference calculation. We can sub-divide the African countries according to their export concentration into three main groups, namely (i) countries depending mainly on agricultural products, (ii) countries depending mainly on natural resources, and (iii) more diversified exporters. A country is allocated either to group (i) or (ii) if the value of its export share in that category exceeds 50 percent in 2009. Otherwise it is allocated to group (iii). Figure 3.5a and Figure 3.5b present the more diversified exporters. None of these exhibit structural breaks, with the exception of South Africa and Mauritius (although we do find structural breaks for Egypt for lower time differences, see Table 3.5). Figure 3.5c shows that 3 out of 5 agricultural product dependent countries (Kenya, Malawi and Cote d' Ivoire) did not show structural change in the given period. Figure 3.5d, in contrast, shows that most natural-resource dependent countries (Algeria, Angola, Libya, Mozambique and Gabon) *did* experience structural change (while Zambia experience structural change for lower time differences, see Table 3.5). More than half of the African countries thus never experienced structural breaks, while countries that are highly dependent on natural-resource exports tend to change their export structure more frequently.

[38] The average number of changes is calculated from the table in Figure 3.3. Take 5-year differences as example; the total number of breaks is 18 for 14 transition periods; the average number of changes is thus 1.29 divided by 20 African countries is 6.43 percent structural changes for 5-year differences.

Figure 3.5: HWM index of Africa by exports concentration[39]



Note: The difference is backward in time; significance at 10% level; 5-year difference, 3-year moving average.

Table 3.5 lists all significant changes in African countries for 1-5 year differences.[40] We cluster the changes into three groups according to the number of changes. Only five countries have more than two structural breaks. We also list the exact years of the structural breaks for ease of reference when we discuss possible causes for these breaks in the next subsection.

---

[39] A country is classified as 'agricultural export concentrated' if the value of its agricultural exports exceed 50% of its total export value in 2009. Similarly for 'natural-resource export concentrated'. The other countries are classified as 'diversified exporters'. The classification of export type is listed in the next subsection.

[40] There is no significant structural change for 1-year differences.

Table 3.5: All significant structural breaks: transition period, 1-5 years difference

Number	Transition	Differences	Transition	Differences	Transition	Differences
1	Libya		Mozambique		Zambia	
	1999, 2004	5	1995, 2000	5	1996, 1999	3
2	Algeria		Egypt		Ethiopia	
	1995, 1999	4	2006, 2008	2	2002, 2006	4
	1995, 2000	5	2006, 2009	3	2002, 2007	5
≥ 3	Angola		Gabon		Mauritius	
	1996, 2001	5	2002, 2007	5	1997, 2001	4
	1997, 2001	4	2003, 2006	3	1997, 2002	5
	1997, 2002	5	2003, 2007	4	1998, 2002	4
	1998, 2003	5	2003, 2008	5	1998, 2003	5
	South Africa		Uganda		1999, 2002	3
	1993, 1996	3	1998, 2003	5	1999, 2003	4
	1993, 1997	4	1999, 2003	4	1999, 2004	5
	1993, 1998	5	1999, 2004	5	2000, 2003	3
	1994, 1997	3	2000, 2003	3	2000, 2004	4
	1994, 1998	4	2000, 2004	4	2000, 2005	5
	1994, 1999	5	2000, 2005	5		

3.5.3 Explanations for certain structural breaks

Although there is no evidence of structural change in African exports if we take all the 20 African countries as a whole, we do find some significant breaks in certain countries (see Table 3.5). Do these breaks really identify structural change or are they driven by data issues? In this subsection, we investigate some possible reasons for these breaks, focusing on the countries that experienced two or more statistically significant breaks.

A. Data issues: Egypt and South Africa

First, we attempt to rule out the observed structural change caused by data issues. As a matter of fact, we do find some data issues in Egypt and South Africa, which have a high fraction of the non-classified commodity (SITC 93) in their export structures in certain years.[41] This commodity is for goods that lack a well-defined classification. A large share of non-classified commodity implies a distorted low share of other commodities of interest, leading to biased Balassa indices for other sectors. In Egypt, the share of the non-classified commodity (SITC 93) surged from 3.45% in 1999 to its peak of 13.4% in 2006, corresponding to the beginning year of identified structural change. It dropped to 4.2% and

[41] SITC 93: Special transactions, commodity not classified according to class

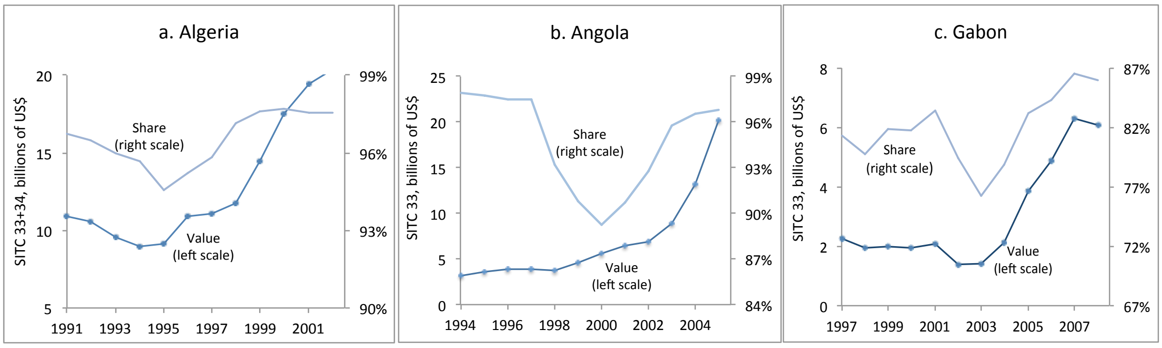
0.2% in 2008 and 2009, respectively, when the identified structural changes ended. Similar developments in South Africa are even more pronounced as the share of the non-classified commodity SITC 93 accounted for 38.5% and 33.5% of total export value in 1993 and 1994, respectively. Again this coincides with identified structural change. We think it is fair to conclude that in all these cases there is no real structural change in these countries, just a classification problem.

B. Pseudo structural change: Libya, Algeria, Angola, Gabon and Ethiopia

We find that some countries indeed show patterns of structural change; however, these changes are temporary and caused by exogenous shocks (such as the outbreak of civil wars, the change of oil prices, and so on). Thus these changes are pseudo structural changes in exports. Usually, these countries are highly relying on one or two types of export products. Taking export values in other sectors as given, a relatively small change in the main export product has a large influence on the total value of exports. The Balassa indices for all other products are affected accordingly, indentified as structural change but easily reversed subsequently. We discuss these countries with ‘pseudo structural change’ separately below.

Libya, Algeria, Angola, and Gabon are all highly dependent on exporting natural resources. Algeria’s exports, for example, are highly concentrated in *Petroleum, petroleum product and related materials* (SITC 33) and *Gas, natural and manufactured* (SITC 34). The share of the two sectors accounted for more than 90 percent of total export values during the sample period (see Figure 3.6a). Thus, a slight change in the exports of the two sectors would severely affect the total export value. By comparing the distributions of the Balassa indices in 1995, 1999 and 2000 for Algeria, we find that Balassa indices in most sectors of 1995 are higher than that of 1999 and 2000. We attribute this phenomenon to the changes of the total exports in SITC 33 and SITC 34 in 1995, 1999 and 2000. In 1995, a dramatic decline in export values of SITC 33 and SITC 34 led to a sudden decline of total value of Algerian exports which therefore increased the Balassa indices of the other sectors. On the contrary, a sudden increase in exports for SITC 33 and 34 in 1999 and 2000 led to a large value of total exports and low Balassa indices for the other sectors. The unbalanced distributions of Balassa indices contribute to the significant breaks in 1995-1999 and 1995-2000 transitions in Algeria.

Figure 3.6: Natural-resource export concentration: the value and share



Source: UN Comtrade

A similar story applies to Angola and Gabon as well. The breakout of the Angola civil war in 1996 hindered the export of *Petroleum, petroleum product and related materials* (SITC 33), the most important export item for Angola which accounted for more than 90 percent of total exports in most of the years (Figure 3.6b). The relatively small exports of SITC 33 during the chaos period (1996-1998) resulted in larger Balassa indices of other sectors compared to the period 2001-2003, leading to the breaks shown in Table 3.5. Similarly for Gabon in 2002-2003 compared to 2006-2008 (Figure 3.6c).

Ethiopia, finally, is a typical African country with a high dependence on agricultural exports, such as *coffee, tea, cocoa, oil seeds and oleaginous fruit, crude animal and vegetable materials*.[42] The year 2002 is special from a structural perspective (Table 3.5). In that year a severe drought ravaged Ethiopia and temporarily damaged its comparative advantage in agricultural products, thus explaining the breaks relative to 2006-2007.

We therefore conclude that the structural changes identified in Libya, Algeria, Angola, Gabon, and Ethiopia are actually consequences of temporary, exogenous shocks, rather than actual structural changes in comparative advantages.

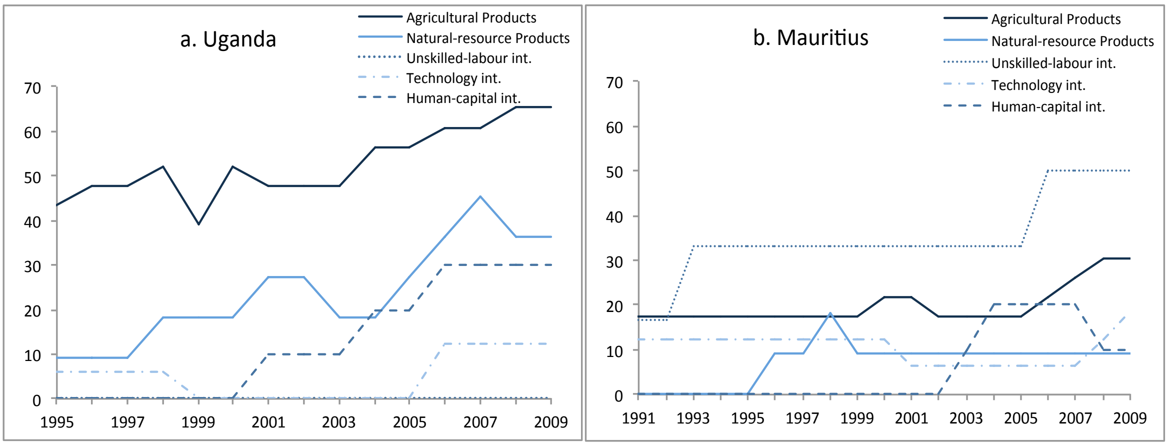
C. Other cases for structural change: Uganda and Mauritius

After discussing data issues and pseudo structural change there are only two countries left with identified structural changes, namely Uganda and Mauritius. To explore the origins of structural change in these countries, we look at their export structure. According to classification of 2-digit SITC commodity in the last subsection, we calculate the share of

[42] SITC 07, SITC 22 and SITC 29

sectors in each product group that has a Balassa index greater than unity (see Figure 3.7). Changes in these shares are expected to reflect changes in trade patterns. The intuition is as follows. The structural changes in comparative advantage (Balassa index) are expected to reflect the dynamics of the domestic industrial structure, thereby reflecting the changes in the export share of different sectors. Figure 3.7 shows the share of sectors in different product groups with Balassa index greater than unity in Uganda and Mauritius.

Figure 3.7: Share of sectors in product group with BI>1, 2-digit



Let us first look at Uganda. Although the development of agriculture was given priority in Uganda since 1986 the involvement in the civil war of the Democratic Republic of Congo from 1998 to 2000 severely impeded Uganda's economic growth. At the same time, the shares of both the agricultural and technology intensive products declined (see Figure 3.7a). Since the recovery period from 2003, the number of sectors with BI>1 increased significantly, exceeding that of 1998, 1999 and 2000. This contributes to the significant breaks shown in Figure 3.5. However, agricultural products have continuously ranked as the most important export sectors in Uganda. In this sense, there is no industry-upgrading structural change from an agriculture-dominant to an industry-dominant one.

Manutius is more complicated. Sugar was its only main export product before the 1980s. With the assistance of the World Bank and the International Monetary Fund (IMF), Mauritius started to diversify its export structure, targeting unskilled-labor intensive product exports. At the beginning of the 1990s, these became the most competitive exports (see Figure 3.7b), replacing agricultural primary products. As an indication of the further diversification in export structure, the share of agricultural primary products and subsequently unskilled-labor products was reduced. The share of sugar (SITC 06) exports, for example, decreased from 24 percent in 1997 to 16.5 percent in 2005. The share of clothing (SITC 84) exports decreased from 56 percent in 1997 to 38 percent in 2005. Many other sectors have been rapidly growing since the new millennium. Although some of them have not reached a comparative advantage yet (their BIs<1, unobservable in Figure 3.7b), we do observe rising Balassa indices, which contribute to the significant breaks reported in Table 3.5. In view of the above, Mauritius might be the only candidate for true structural change in Africa. According to the African Economic Outlook (2014, African Development Bank), Mauritius has been benefiting from its strong institutions. The stable and less corrupted government has been encouraging the foreign investment and sustaining structural reform, which help Mauritius succeed.

3.5.4 Markov transition and mobility indices

The Harmonic Weighted Mass (HWM) index analysis in the above two subsections suggests that most African countries did not experience structural change in their export profile over the past two decades. This section investigates transition probability matrices and mobility indices to analyze the degree of persistence and mobility of revealed comparative advantage (Balassa index). Our main objective is to corroborate our earlier findings from a different perspective.

To identify whether the mobility of African export structure is large or not, we use the results for the EU-12 in Hinloopen and van Marrewijk (2001) and for the G-5 in Proudman and Redding (2000) as benchmarks since their data sets are similar to ours.[43] We also use the same classification rules. This leads to four different states in both cases, namely No, Weak, Medium and Strong comparative advantage as in Hinloopen and van Marrewijk (2001), see Table 3.6a, and the four quartiles as in Proudman and Redding (2000), see Table 3.6b. In the latter case the Balassa indices are sorted from low to high in each year for all countries (pooled sample) and for individual countries. *Quartile1* refers to the lowest ranked 25 percent in all countries or in each African country, and so on.

---

[43] 2-digit Combined Nomenclature Industry code (99 sectors) from Eurostat is used in Hinloopen and van Marrewijk (2001). The data of 22 industries classified from ISIC is used in Proudman and Redding (2000). In our paper, we use 2-digit SITC 2 data (69 sectors) from UN Comtrade.

Table 3.6: Categories of the Balassa index

a. Category I: No, low, medium, high (Hinloopen and van Marrewijk (2001))		
State 1	0 < Balassa index ≤ 1	No Comparative Advantage
State 2	1 < Balassa index ≤ 2	Weak Comparative Advantage
State 3	2 < Balassa index ≤ 4	Medium Comparative Advantage
State 4	4 < Balassa index	Strong Comparative Advantage
b. Category II: Quartiles (Proudman and Redding (2000))		
State 1	0% < Balassa index ≤ 25%	Quartile 1
State 2	25% < Balassa index ≤ 50%	Quartile 2
State 3	50% < Balassa index ≤ 75%	Quartile 3
State 4	75% < Balassa index ≤100%	Quartile 4

Assuming that the stochastic process determining the evolution of revealed comparative advantage (Balassa index) is the same across each economy, we analyze a pooled sample for all African countries as a whole.[44] Table 3.7 shows the results of Markov transition probabilities according to the two classifications for a 1-year transition. Each row in the sub-table denotes the estimated probability of passing from one state to another. Generally, the larger the diagonal data the greater the probability of persistence in each state, while the larger the off-diagonal terms the greater mobility. For example, the first data in Table 3.7a denotes that the sectors without comparative advantage have 97.7 percent probability of remaining in the no comparative advantage state in the next year. The other data in the first row presents the probabilities of moving into the state of the weak, medium and strong comparative advantage in the next period, respectively.

The last two rows of Table 3.7 present the Ergodic distribution for each state in two categories. The theoretical distribution reflects the stationary distribution of the revealed comparative advantage of Africa after a large number of transitions. The empiric distribution corresponds to the distribution of the Balassa indices during the entire sample period. For example, due to the quartile classification in category II, the empirical *ergodic* distributions are 25 percent in each state. However, the stationary distribution changes substantially, especially the share of Balassa indices in states 1 and 4. The share of the Blassa indices in state 1 decreases from 25 percent in sample period to 17.5 percent in its

[44] Four countries in the sample, listed in Appendix 3.5, are dropped in the calculation of Markov transition probability matrix and mobility index for various reasons. There is only one observation (SITC 61 with Balassa index 3.34 in 2000) for Mali in state 3 for Category I. This unbalanced distribution results in movement from state 3 to another state with 100 percent probability (see Appendix 3.5-Mali). There are no ergodic distributions in the other three countries (Algeria, Libya and Mauritius) since their transition matrices are not irreducible.

stationary status, while the distribution in state 4 changes from 25 to 35.3 percent. To relax the assumption in the beginning of the last paragraph, we also estimate the transition probability matrix for individual African countries (see Appendix 3.6-Appendix 3.8).

Table 3.7: Markov 1-year transition matrices for Africa by two categories: the pooled sample

a. Category I: No, weak, medium, strong					b. Category II: Quartiles				
1-year	No	Weak	Medium	Strong	1-year	1	2	3	4
No	0.977	0.020	0.002	0.000	1	0.898	0.097	0.004	0.001
Weak	0.151	0.726	0.118	0.005	2	0.084	0.821	0.093	0.002
Medium	0.008	0.117	0.768	0.107	3	0.003	0.068	0.866	0.063
Strong	0.001	0.004	0.061	0.934	4	0.001	0.002	0.047	0.950
Theory	0.644	0.090	0.097	0.169	Theory	0.175	0.202	0.270	0.353
Empiric	0.749	0.080	0.069	0.102	Empiric	0.250	0.250	0.250	0.250

Note: Some rows in the matrix do not add up to one due to rounding, as in Appendix 3.5-3.8; the theoretical ergodic distribution is the limiting distribution; the empirical one is the number of observations in that state for the period as a whole.

A mobility index is calculated on the basis of the transition probability matrix and indicates the degree of mobility in the Balassa index distribution. Table 3.8 shows the mobility indices for the pooled samples for the benchmark and the sixteen African countries. In the pooled sample, all the mobility indices for Africa are lower for the benchmark in both categories in 1-year transitions, especially in category I. In other words, the changes in the distribution of the revealed comparative advantage (Balassa index) are fewer in African countries than in the EU-12 and G-5, respectively, taking all countries in each sample as a whole.

Table 3.8: Mobility indices for pooled sample: 1-year transition

	Category I: No, weak, medium, strong				Category II: Quartiles		
Country	M1	M2	M3	M4	M1	M2	M3
Benchmark	0.330	0.180	0.730	0.120	0.163	0.121	0.426
Africa-16	0.198	0.161	0.510	0.032	0.155	0.115	0.407

Note: The benchmark in Category I is for 12 EU countries (Hinloopen and van Marrewijk, 2001); the benchmark for Category II is for the G-5 (Proudman and Redding, 2000).

We also calculate the mobility indices M1-M4 for individual African countries.[45] In order to see whether these mobility indices identify similar patterns of mobility for individual countries, we check the correlations of these mobility indices within each category and

[45] We only calculate M1-M3 for category II since there is no M4 in Proudman and Redding (2000).

between the two categories, see Figure 3.8 and Table 3.9. There is a clear indication that the correlations are higher within the category than between the two categories. In sum, all the correlation diagrams and the correlation matrix show that these indices are highly positively correlated, confirming that all these indices are capable of measuring mobility.

Figure 3.8: Correlations of M1-M3 between category I and category II

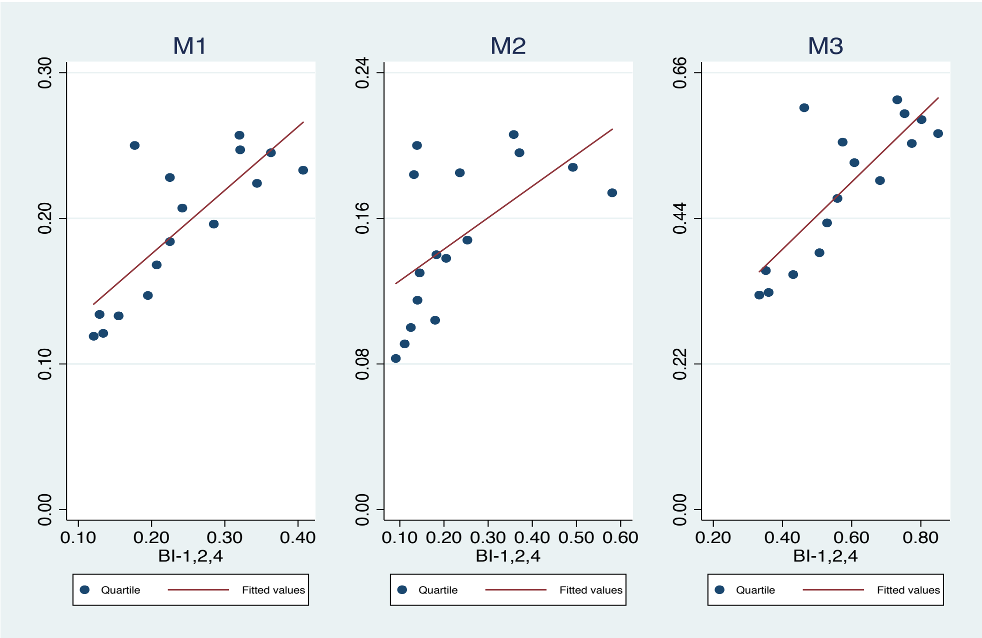
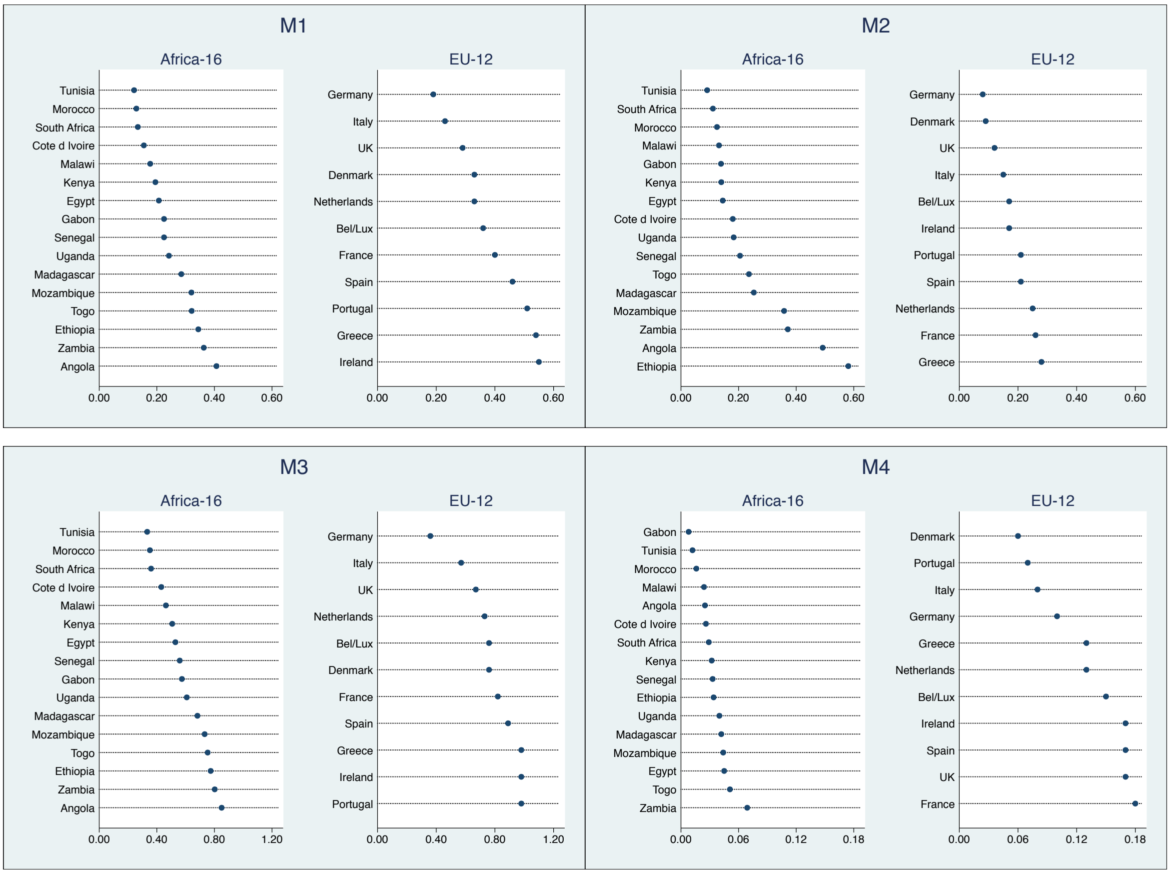


Table 3.9: Correlation matrix of mobility index

	BI-M1	BI-M2	BI-M3	BI-M4	Quartile-M1	Quartile-M2	Quartile-M3
BI-M1	1.000						
BI-M2	0.870	1.000					
BI-M3	0.994	0.84	1.000				
BI-M4	0.573	0.37	0.601	1.000			
Quartile-M1	0.785	0.58	0.808	0.432	1.000		
Quartile-M2	0.791	0.59	0.810	0.368	0.976	1.000	
Quartile-M3	0.784	0.58	0.808	0.434	0.999	0.969	1.000

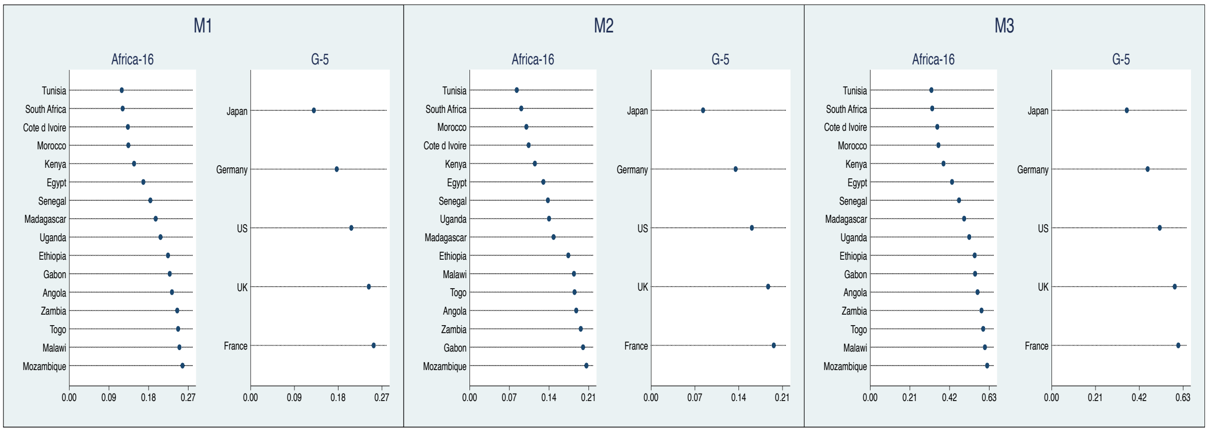
After checking the correlations of all the mobility indices, we compare each mobility index with its corresponding benchmark. Figure 3.9 shows the M1-M4 of each African country and benchmarks in category I (see Appendix 3.9 for details). The longer the distance between the dot and the vertical axis, the greater mobility is. It is clear that the mobility indices for African countries are lower than for the EU-12 in the 1-year transition, except for M2. In other words, there is little evidence of change in revealed comparative advantage for African exports.

Figure 3.9: M1-M4 for individual African countries and benchmarks, category I



Regarding category II (see Figure 3.10), the three mobility indices demonstrate a similar pattern for Africa and the G-5 as for the pooled sample (see Table 3.8). We attribute it to the classification of quartiles in category II. The quartile division of the set of Balassa indices generates different state classifiactions between our sample and the benchmark. In other words, the boundaries between two states for Africa are different from that for the G-5, in contrast to the fixed boundaries in category I. As a result we do not think the different states between our sample and the benchmark can be adequately compared and expect that there is no big gap between the mobility indices under the classification of quartiles, as shown in the comparison of Africa and the G-5 in both pooled sample and individual countries in category II (see Table 3.8 and Figure 3.10).

Figure 3.10: M1-M3 for individual African countries and benchmarks, category II



To summarize, by comparing mobility indices with the corresponding benchmark in the pooled sample and for individual countries, we conclude that the mobility of revealed comparative advantage in African export trade is *lower* than for the benchmarks (EU-12 and G-5) in the sample periods. The lower mobility supports the earlier finding of the absence of structural change in Africa for its exports from 1991 to 2009.

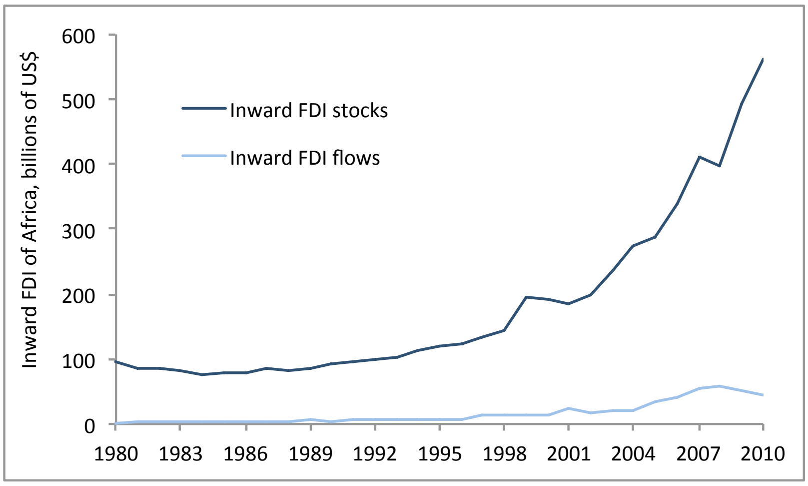
3.6 Possible explanations for the absence of structural change

In section 3.5, by analyzing the evolution of the entire distribution of Balassa indices, we reach the same conclusion based on several indicators; namely, there is no evidence of structural breaks in African exports during the last two decades. In other words, over time, African exports remain concentrated on their primary industries, more specifically the natural-resource intensive and agricultural products, which means that the recent rapid economic growth in Africa is not accompanied by structural breaks and upgrading of export structure. In this section, we explore two main reasons for the lack of structural changes in African exports, namely (i) persistence in exporting primary goods and natural-resource intensive goods and (ii) barriers to switching to manufacturing production.

3.6.1 Persistence in exporting primary and natural-resource intensive goods

Africa has maintained its comparative advantages mainly in primary products over the last two decades (see Table 3.3). We attribute this phenomenon to its abundant endowments in natural resources and land. According to the Heckscher-Ohlin theorem, capital abundant countries export capital-intensive manufactured products in exchange for labor-intensive products from labor abundant countries, and vice versa. The predictions of the Heckscher-Ohlin (HO) factor abundance model have been empirically tested for a long time.[46] Leontief (1953) found that the capital-labor ratio in US imports was higher than in US exports, contradicting the assumption that the US was a capital abundant country. Leamer (1980), however, pointed out that a proper test should be based on the factor content of trade flows (see Vanek (1968)). Early empirical tests of the factor content of trade generally found that the model performed poorly when compared against the data (see for example Leamer (1984) and Bowen et al. (1987)). The availability of better and more detailed data allowed Trefler (1995) to identify the systematic pattern of departures in the data from the predictions of the model (labeled the "mystery of the missing trade").[47] This, in turn, enabled Davis et al. (1997) to identify the importance of non-factor price equality and Davis and Weinstein (2001) to explain the data rather successfully once technical differences, non-factor price equality, non-traded goods, and the costs of trade are taken into consideration. This also holds when trade in intermediate goods is explicitly incorporated in the analysis (see Trefler and Zhu, 2010).

Figure 3.11: Inward FDI stocks and flows in Africa, 1980-2010



Source: UNCTAD stat and WDI. Constant 2010 US$

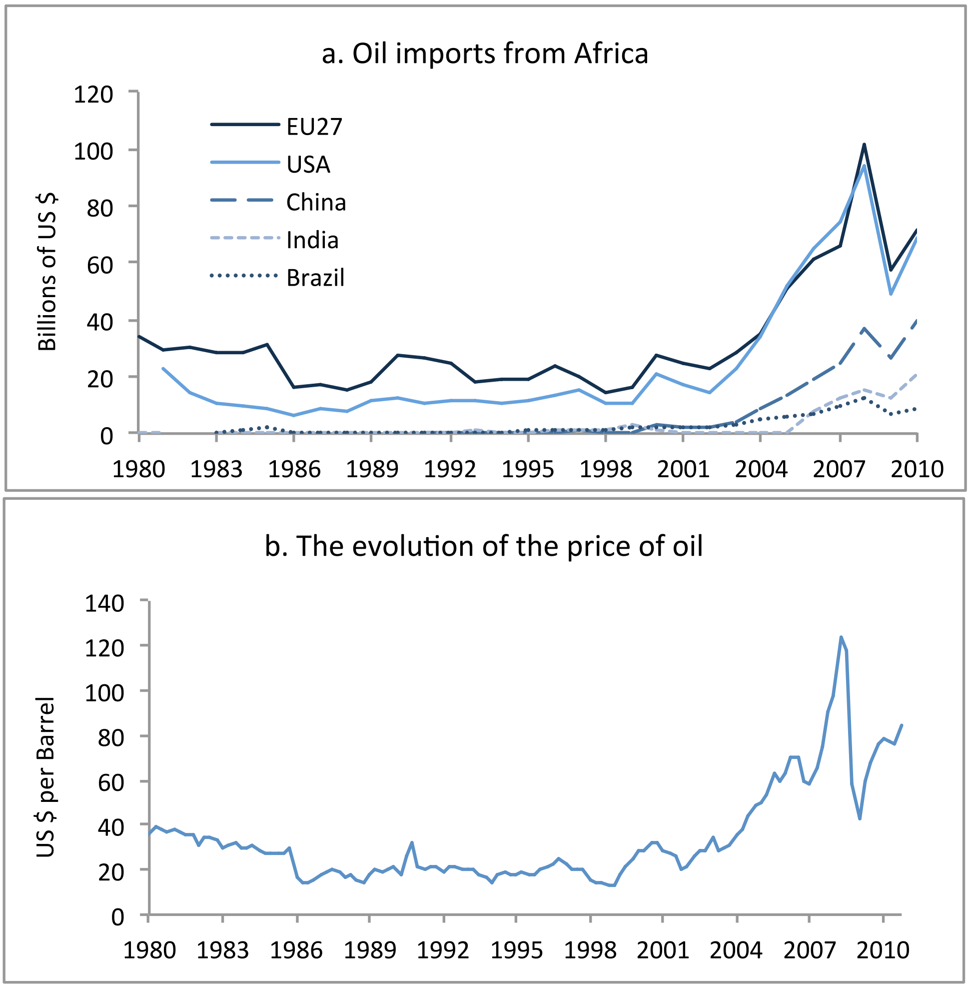
---

[46] See Feenstra (2004), Baldwin (2008), or Leamer (2012) for recent surveys of the literature. The description below is taken from Brakman and van Marrewijk (2013).

[47] The "missing trade puzzle" refers to the fact that the predicted factor content of net exports is smaller than the actual factor content, hence trade is "missing." In addition, two groups of countries can be identified: developing and developed countries. For poor countries the difference between actual and predicted factor content of net exports is negative, while for rich countries this is positive. This implies that poor countries are abundant in most factors of production, whereas rich countries are scarce in most factors of production.

Although natural-resource exports are natural resource intensive, the extraction of natural resources essentially requires capital and knowledge. However, Africa is not a capital and knowledge abundant continent. In fact, along with knowledge, most capital invested in the natural resource products originates from a large number of foreign investors. The natural-resource extraction industry attracts continued foreign investment as a result of large-scale profits in the international market. Figure 3.11 shows the inward FDI stocks and flows in Africa since 1980. It is clear that the trend of inward FDI for Africa was stable in the 1980s. Later on, however, inward FDI rose gradually in the last two decades in both stocks and flows, especially after 2000. The main reason is that on top of the traditional investors from Europe and America increased capital flows were coming from emerging markets, such as China, India, and Brazil.

Figure 3.12: The dynamics of oil imports from Africa and oil prices, 1980-2010



Source: a. Author calculate based on UN Comtrade, oil refers to SITC33; b. Federal Reserve Bank of ST. Louis, see http://research.stlouisfed.org/fred2/series/OILPRICE/downloaddata, Quarterly data.

On the one hand, both traditional and new participants bring large investments that contribute to the natural resource extraction industry. On the other hand, their large demand for natural resources in their domestic markets pushes up the prices of natural resources in the international market, which further encourages the growth of export-oriented natural resource sectors in Africa. Take oil, for example, the most important product of African export. The demand for oil in each region is highly correlated with the oil prices over time. EU-27 and USA were still the primary destinations of African oil exports in the last three decades, but the strong demand for oil came from the emerging economies (China, India and Brazil) since 2000 (see Figure 3.12a). Recovering from the second oil crisis in the beginning of the 1980s, oil prices kept steady during the 1990s. Along with the increasing demand for oil in the 21st century, the oil price has been growing rapidly, except for the period of the recent financial crisis (see Figure 3.12b). In sum, both the foreign investment and increasing demand for oil contribute to the persistence in exporting primary goods in Africa over time.

3.6.2 Barriers to shifting to manufacturing production

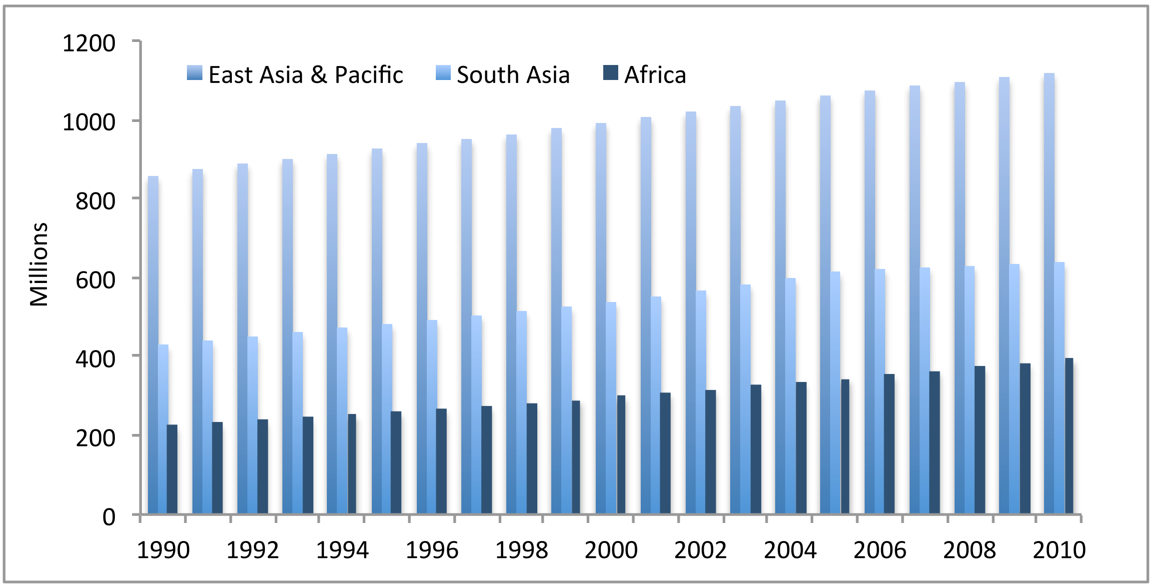
Simple manufacturing industries are most likely to develop first in most countries. This does not seem to be the case, however, in Africa. Though Africa is not a capital abundant continent, it is still possible to develop the export-oriented manufactured industry if it attracts enough foreign direct investment, as do many natural resource industries. Unfortunately, foreign investments are reluctant to shift to the non-resource extraction industries for two reasons. First, these foreign investors benefit from the existing investment in African natural resource industries, as we discussed above. Second, and more importantly, the disadvantages of Africa in terms of labor force, geographic factors and institutions seem to be big barriers for foreign investors to develop manufacturing industries. In this subsection, we analyze the comparative disadvantage of Africa in three perspectives, comparing the data of Africa to that of East Asia & Pacific and South Asia.[48] The two regions used for comparison are chosen as they have a comparative advantage in simple manufactured products.

[48] We refer to the developing countries in East Asia and Pacific.

A. Labor force

First of all, Africa lacks advantage in labor quantity. Though the area of Africa is much larger than that of East Asia & Pacific and South Asia, the African labor force is lower than either that of East Asia & Pacific or that of South Asia in quantity (see Figure 3.13). Historically, East Asia & Pacific has been one of the most populous regions, thus it is not surprising that East Asia & Pacific have the largest labor force in the last two decades among the three regions. On the contrary, Africa is endowed with the least labor force over time due to natural and social reasons, such as large desert areas, civil wars, poor health care, and so on. The relatively low quantity of African labor force is not conducive to develop basic labor-intensive manufacturing industries, such as the textile industry.

Figure 3.13: Labor force, 1990-2010[49]



Source: WDI and ADI from World Bank

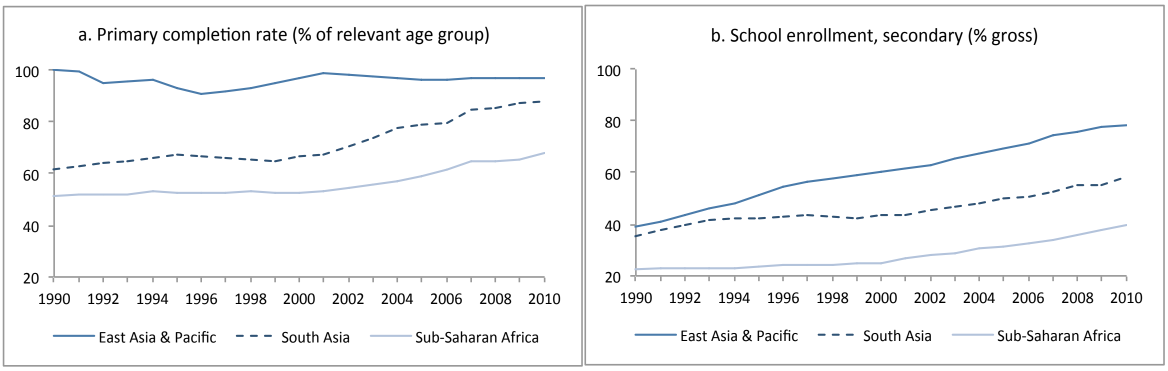
Second, and more importantly, Africa lacks an advantage in labor quality as well. Africa is abundant of labor with low skills, leading to inefficient production even in the simple types of manufactures. The unskilled labor is the result of low education levels in Africa. Figure 3.14 shows two indicators in terms of primary and secondary education in East Asia & Pacific, South Asia, and Sub-Saharan Africa.[50] In both the primary completion rate and secondary school enrollment, the index of Africa is lower than that of the other regions.

[49] The labor force comprises people ages 15 and older.

[50] There is no aggregate data about the entire Africa, thus the data of Sub-Saharan Africa (all income levels) is adopted.

The primary completion rate in Africa is around 50 to 60 percent from 1990 to 2010 (see Figure 3.14a). At the same time, the enrollment of secondary school was increasing from 20 percent in 1990 to about 40 percent in 2010 (see Figure 3.14b). By 2010, more than one half of African people did not receive secondary education. The lack of skilled labor prevents enterprises to invest and build factories in Africa since the training of the labor force may be more costly than that in Asian countries.

Figure 3.14: The education of the labor, 1990-2010



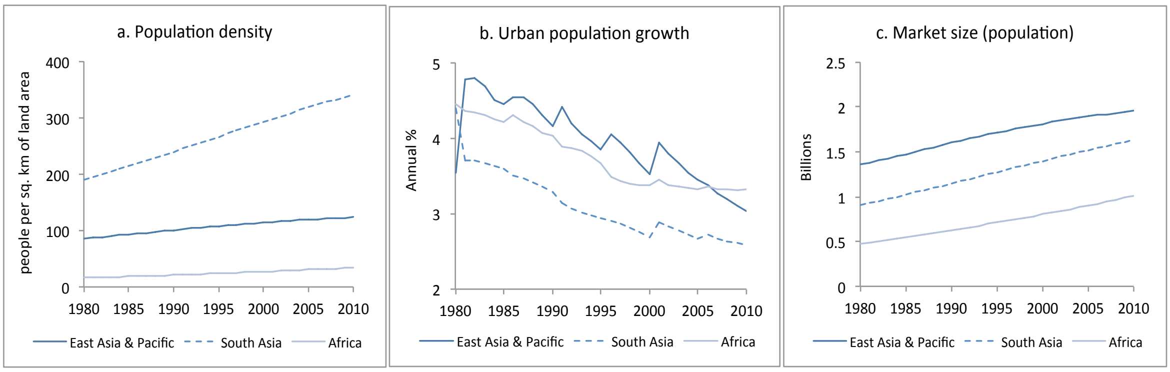
Source: WDI from World Bank

B. Geographic factors

Since the *new economic geography* literature developed at the beginning of 1990s, the location of economic activities in space has attracted a lot of interest. Generally speaking, the most striking feature of the geography of economic activities is *concentration* (Krugman, 1991), leading to density of populations and urban areas. Industries concentrate in space since these urban areas are highly specialized and productive. In addition, a large density of population in the urban areas implies large demand for local manufactures, thus lowering transportation costs. Both the specialized production and the large market size contribute to economies of scale so that the production of many manufactures reinforce itself in these areas.

Africa has a comparative disadvantage in geographic factors. In other words, the diverse geography does not stimulate the concentration required to develop manufacturing activity in Africa. We illustrate this in terms of population density, urbanization, and market size (see Figure 3.15). Compared to the data for East Asia & Pacific and South Asia we note that Africa does not have a comparative advantage in population density or market size over the last three decades. The urbanization process is fastest in East Asia & Pacific and the population is largest as well.[51] The market size is lowest in Africa, certainly if we correct for the degree of urbanization. The lack of good geographic conditions thus impedes the development of manufactures in Africa.

Figure 3.15 Three variants of concentration, 1980-2010



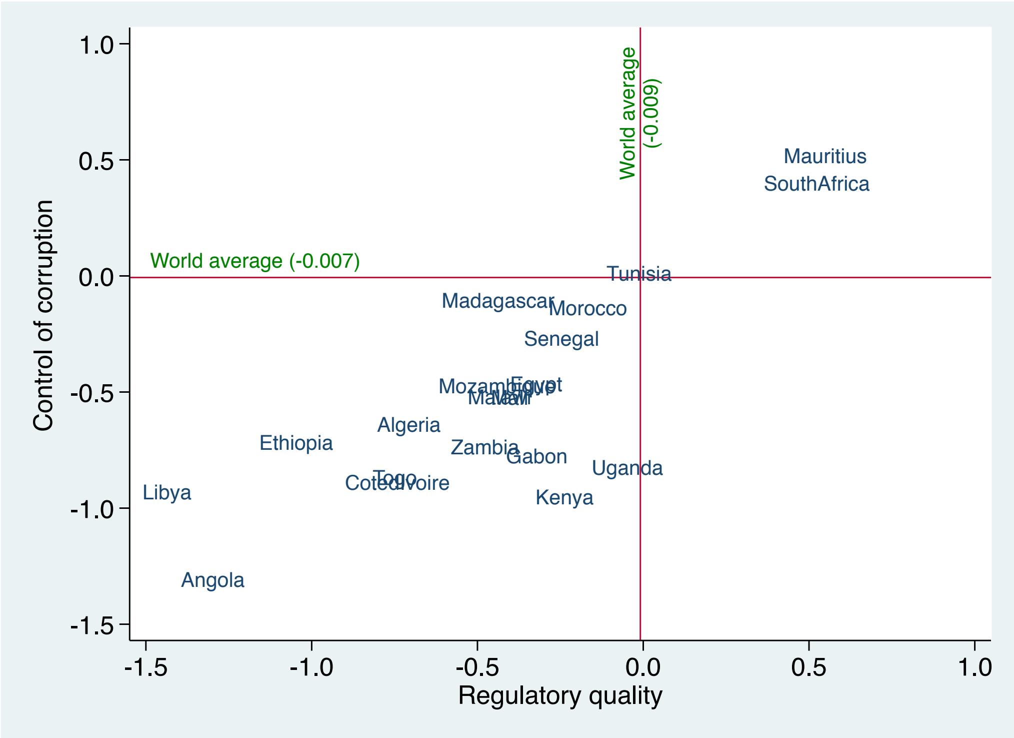
Source: WDI and ADI from World Bank

C. Institutional factors

Good institutions are key to cultivate manufacturing industries (North, 1990). On the one hand, regions with a good institution attract more foreign direct investment than peers with bad institutions. In the case of Africa with insufficient domestic investment, foreign investment is the prerequisite for the development of manufacturing industries and therefore upgrading structural change. On the other hand, bad institutions such as corruption and unstable governments have a negative effect on the regulatory control and the social equality. Africa does not score well on institutional factors, as indicated in Figure 3.16 for two institutional indicators. It is clear that the control of corruption and regulatory quality of most Africa countries in our sample is lower than the world average, especially for the oil rich countries Angola and Libya. The two outliers in this respect are Mauritius and South Africa. The bad institutions give rise to severe inequality within countries. The benefits of natural resource extraction may thus be limited to a few groups of people, who suppress other people to consolidate these benefits. The natural resource industry reinforces itself overtime, reducing the opportunity to develop other manufactures.

[51] The urbanization process in East Asia & Pacific and South Asia may be underestimated by the official data used in Figure 3.15b since a large number of labor-intensive manufactures in urban areas attract laborers from rural areas. These people live and work in urban areas, but they are excluded from the official data analyzed.

Figure 3.16: Institutional indicators of Africa (average of 1996-2010)



Source: The worldwide government indicators[52], www.govindicators.org

In sum, both the natural resource endowments and the foreign investment in natural resource extraction contribute to the continuous development of natural resource industry in Africa. At the same time, the comparative disadvantages of labor endowments, geographic factors and institutions prevent foreign investment to shift from traditional natural resource producing industry to other manufactures. All these activities jointly lead to the absence of upgrading structural change in Africa. Some African countries realize the importance of developing manufacturing industries and are trying to attract foreign direct investment through preferential policies. Similarly, some advanced regions are trying to help African develop manufacturing industries. The *Africa Growth and Opportunity Act* (AGOA), for example, was established by the USA and European countries and uses tariff reduction and exemption specifically for the imports from Africa. Along with the increase of labor quantity and quality in Africa, the rising wage rates in East Asia, some basic manufacture industries are moving to Africa, making it possible that we observe structural breaks in African exports in the future.

---

[52] These indicators range from approximately -2.5 (weak) to 2.5 (strong) governance performance.

3.7 Conclusions

In this study, we empirically identify the absence of structural breaks in African exports using a sample of 20 African countries over the last two decades. Our analysis focuses on changes in revealed comparative advantage using the Balassa index as an indicator. First, we use the HWM index to identify structural breaks and we hardly find any significant structural breaks at all for the African countries. The few structural breaks we do find for some countries are mostly caused by data issues and pseudo structural changes. Second, we turn to alternative tools to investigate the same question. We calculate Markov transition matrices and mobility indices to analyze persistence and mobility of revealed comparative advantage. We generally find that persistence is high and mobility is low in African countries. Both findings indicate that although the African countries have been developing rapidly in the recent past this development is (not yet) accompanied by the type of structural changes we observe in other countries. We briefly discuss some potential underlying causes for the observed absence of structural change. First, the strong advantages in exporting primary and natural-resource intensive goods from Africa leads to a persistent lock-in that prevents the development of other sectors. Second, there are several obstacles to switch to producing basic manufactures, such as the size and quality of the labor force, geographic factors (density, urbanization, and market size), and institutional factors. In light of this, the continued growth of the African population, accompanied by a further increase in urbanization may create the required concentration, density, and market size, particularly in combination with the rising wages in other parts of the world, to make it possible for the manufacturing process to take off in the near future. A question to be answered only by future analysis.

3.8 References

Ariovich, G. (1979). The comparative advantage of South Africa as revealed by export shares. *South African Journal of Economics* 47(2), 123-129.

Balassa, B. (1965). Trade liberalization and 'revealed' comparative advantage, *The Manchester School of Economic and Social Studies* 33, 92-123.

Balassa, B. (1977). Revealed'comparative advantage revisited: An analysis of relative export shares of the industrial countries, 1953–1971. *The Manchester School* 45(4), 327-344.

Balassa, B. (1979). The changing pattern of comparative advantage in manufactured goods. *The Review of Economics and Statistics* 61(2), 259-266.

Baldwin, R.E. (2008). *The development and testing of Heckscher–Ohlin trade models*. MIT Press, Cambridge, Mass.

Bartholomew, D.J. (1973). *Stochastic models for social processes*, second edition, Wiley, London.

Bowen, H.P., Leamer, E., and Sveikauskus, L. (1987). Multi-country, multifactor tests of factor abundance theory. *American Economic Review* 77, 791–809.

Brakman, S., Inklaar, R., and Van Marrewijk, C. (2013). Structural change in OECD comparative advantage. *The Journal of International Trade and Economic Development*, 22(6): 817-838.

Brakman, S. and C. van Marrewijk (2013). Lumpy countries, urbanization, and trade. *Journal of International Economics* 89(1): 252-261.

Brasili, A., Epifani, P., and Helg, R. (2000). On the dynamics of trade patterns. *De Economist* 148(2), 233-258.

Dalum, B., Laursen, K., and Villumsen, G. (1998). Structural change in OECD export specialisation patterns: de-specialisation and 'stickiness'. *International Review of Applied Economics* 12(3), 423-443.

Davis, D.R., Weinstein, D.E., Bradford, S.C., and Shimpo, K. (1997). Using international and Japanese regional data to determine when the factor abundance theory of trade works. *American Economic Review* 87 (3), 421–446.

Davis, D.R., and Weinstein, D.E. (2001). An account of global factor trade. *American Economic Review* 92, 1423–1453.

De Benedictis, L., and Tamberi, M. (2001). A note on the Balassa index of revealed comparative advantages, *SSRN Working Papers* (http://ssrn.com/abstract=289602).

Edwards, L., and Schoer, V. (2002). Measures of competitiveness: a dynamic approach to South Africa's trade performance in the 1990s. *South African Journal of Economics* 70(6), 1008-1046.

Feenstra, R. (2004). *Advanced international trade: Theory and evidence*. Princeton University Press, Princeton.

Grossman, G. M., and Helpman, E. (1991). *Innovation and growth in the global economy*. The MIT Press.

Hillman, A.L. (1980). Observations on the relation between 'revealed comparative advantage' and comparative advantage as indicated by pre-trade relative prices", *Weltwirtschaftliches Archiv* 116: 315 – 321.

Hinloopen, J., van Marrewijk, C. (2001). On the empirical distribution of the Balassa index, *Review of World Economics* 137, No. 1, 1-35

Hinloopen, J., and van Marrewijk, C. (2004). Dynamics of Chinese comparative advantage. *Tinbergen Institute Discussion Paper* 04-034/2.

Hinloopen, J., and van Marrewijk, C. (2005). Comparing distributions: The harmonic mass index. *Tinbergen Institute Discussion Paper*, TI 2005-122/1, Rotterdam.

Hinloopen, J., Wagenvoort, R., and van Marrewijk, C. (2012). A K-sample homogeneity test: the Harmonic Weighted Mass index. *International Econometric Review* 4(1), 17-39

Krugman, P. R. (1991). *Geography and trade*. MIT press.

Laursen, K. (1998). Revealed comparative advantage and the alternatives as measures of international specialisation (No. 98-30). DRUID, Copenhagen Business School, Department of Industrial Economics and Strategy/Aalborg University, Department of Business Studies.

Leamer, E.E. (1980). The Leontief paradox reconsidered. *Journal of Political Economy* 88, 495–503.

Leamer, E.E. (1984). *Sources of comparative advantage: Theory and evidence*. MIT Press, Cambridge, Mass.

Leamer, E.E. (2012). *The craft of economics: Lessons from the Heckscher–Ohlin framework*. The MIT Press, Cambridge, Mass.

Leontief, W. (1953). Domestic production and foreign trade; the American capital position re-examined. *Proceedings of the American philosophical Society*, 332-349.

Lucas Jr, R. E. (1988). On the mechanics of economic development. *Journal of monetary economics* 22(1), 3-42.

Ng, F., Yeats, A.J. (2001). The recent trade performance of Sub-Saharan African countries: Cause for hope or more of the same? *Africa Region Working Paper Series,* No. 26, June, World Bank, Washington, DC.

North, D. C. (1990). *Institutions, Institutional Change and Economic Performance*. Cambridge University Press, Cambridge.

Proudman, J., and Redding, S. (2000). Evolving patterns of international trade. *Review of International Economics* 8 (3), 373-396.

Quah, D. (1993). Empirical cross-section dynamics in economic growth, *European Economic Review* 37, 426-34

Ross, S. M. (2000). *Introduction to probability models*, seventy edition, Academic Press, New York

Shorrocks, A. (1978). The measurement of mobility. *Econometrica* 46, 1013-1024

Sommers, P.S., and J. Conlisk (1979). Eigenvalue immobility measures for Markov chains. *Journal of Mathematical Sociology* 6: 253-276.

Trefler, D. (1995). The case of the missing trade and other mysteries. *American Economic Review* 85, 1029–1046.

Trefler, D., and Zhu, S.C. (2010). The structure of factor content predictions. *Journal of International Economics* 82, 195–207.

United Nations (2011). Department of Economic and Social Affairs, Population Division. World Population Prospects: The 2010 Revision, CD-ROM

Vanek, J. (1968). The factor proportions theory: the N-factor case. *Kyklos* 21, 749–754.

Wood, A., and Berge, K. (1997). Exporting manufactures: human resources, natural resources, and trade policy. *The Journal of Development Studies* 34(1), 35-59.

Wood, A., and Mayer, J. (2001). Africa's export structure in a comparative perspective. *Cambridge Journal of Economics*, *25*(3), 369-394.

3.9 Appendix

Appendix 3.1: 20 African countries

Algeria	Gabon	Mali	South Africa
Angola	Kenya	Mauritius	Togo
Cote d'Ivoire	Libya	Morocco	Tunisia
Egypt, Arab Rep.	Madagascar	Mozambique	Uganda
Ethiopia	Malawi	Senegal	Zambia

Appendix 3.2: 181 Reference countries

Afghanistan	Comoros	Iceland	Micronesia	Slovenia
Albania	Congo, Rep.	India	Moldova	Solomon Islands
Algeria	Costa Rica	Indonesia	Mongolia	South Africa
Angola	Cote d'Ivoire	Iran, Islamic Rep.	Morocco	Spain
Antigua & Barbuda	Croatia	Iraq	Mozambique	Sri Lanka
Argentina	Cyprus	Ireland	Myanmar	St. Kitts and Nevis
Armenia	Czech Rep.	Israel	Namibia	St. Lucia
Australia	Denmark	Italy	Nepal	St. Vincent and the Grenadines
Austria	Djibouti	Jamaica	Netherlands	Sudan
Azerbaijan	Dominica	Japan	New Zealand	Suriname
Bahamas	Dominican Rep.	Jordan	Nicaragua	Swaziland
Bahrain	Ecuador	Kazakhstan	Niger	Sweden
Bangladesh	Egypt, Arab Rep.	Kenya	Nigeria	Switzerland
Barbados	El Salvador	Kiribati	Norway	Syrian Arab Republic
Belarus	Equatorial Guinea	Korea, Rep.	Oman	Tajikistan
Belgium	Eritrea	Kuwait	Pakistan	Tanzania
Belize	Estonia	Kyrgyz Republic	Palau	Thailand
Benin	Ethiopia	Lao PDR	Panama	Timor-Leste
Bhutan	Fiji	Latvia	Papua New Guinea	Togo
Bolivia	Finland	Lebanon	Paraguay	Tonga
Bosnia & Herzegovina	France	Lesotho	Peru	Trinidad and Tobago
Botswana	Gabon	Liberia	Philippines	Tunisia
Brazil	Gambia	Libya	Poland	Turkey
Brunei Darussalam	Georgia	Lithuania	Portugal	Turkmenistan
Bulgaria	Germany	Luxembourg	Qatar	Uganda
Burkina Faso	Ghana	Macao SAR, China	Romania	Ukraine
Burundi	Greece	Macedonia, FYR	Russian Federation	United Arab Emirates
Cambodia	Grenada	Madagascar	Rwanda	United Kingdom
Cameroon	Guatemala	Malawi	Samoa	United States
Canada	Guinea	Malaysia	Sao Tome and Principe	Uruguay
Cape Verde	Guinea-Bissau	Maldives	Saudi Arabia	Uzbekistan
Central African Rep.	Guyana	Mali	Senegal	Vanuatu
Chad	Haiti	Malta	Seychelles	Venezuela, RB
Chile	Honduras	Mauritania	Sierra Leone	Vietnam
China	Hong Kong SAR, China	Mauritius	Singapore	Yemen, Rep.
Colombia	Hungary	Mexico	Slovak Republic	Zambia & Zimbabwe

Appendix 3.3: Classification of 2-digit sectors[53]

A: Agricultural primary products *(23 sectors)*	
00	Live animals chiefly for food
01	Meat and preparations
02	Dairy products and birds' eggs
03	Fish, crustacean and molluscs, and preparations thereof
04	Cereals and cereal preparations
05	Vegetables and fruit
06	Sugar, sugar preparations and honey
07	Coffee, tea, cocoa, spices, and manufactures thereof
08	Feeding stuff for animals (not including unmilled cereals)
09	Miscellaneous edible products and preparations
11	Beverages
12	Tobacco and tobacco manufactures
21	Hides, skins and furskins, raw
22	Oil seeds and oleaginous fruit
23	Crude rubber (including synthetic and reclaimed)
24	Cork and wood
25	Pulp and waste paper
26	Textile fibres (not wool tops) and their wastes (not in yarn)
29	Crude animal and vegetable materials, nes
41	Animal oils and fats
42	Fixed vegetable oils and fats
43	Animal and vegetable oils and fats, processed, and waxes
94	Animals, live, nes, (including zoo animals, pets, insects, etc)
B: Natural resource and natural-resource intensive products (*11 sectors*)	
27	Crude fertilizer and crude minerals
28	Metalliferous ores and metal scrap
32	Coal, coke and briquettes
33	Petroleum, petroleum products and related materials
34	Gas, natural and manufactured
35	Electric current
61	Leather, leather manufactures, nes, and dressed furskins
63	Cork and wood, cork manufactures
66	Non-metallic mineral manufactures, nes
68	Non-ferrous metals
97	Gold, non-monetary (excluding gold ores and concentrates)
C: Unskilled-labor intensive products (*6 sectors*)	
65	Textile yarn, fabrics, made-up articles, nes, and related products
81	Sanitary, plumbing, heating, lighting fixtures and fittings, nes
82	Furniture and parts thereof
83	Travel goods, handbags and similar containers
84	Articles of apparel and clothing accessories
85	Footwear

[53] The classification of 2-digit sectors SITC 52, 66, 67, 76, 79, 88, 89 is ambiguous because their (3-digit) subsectors belong to different groups in http://www2.econ.uu.nl/users/marrewijk/eta/index.htm. We classify them in the group where most of their subsectors belong to.

Appendix 3.4: Classification of 2-digit sectors (continued)

D: Technology intensive products (*16 sectors*)	
51	Organic chemicals
52	Inorganic chemicals
54	Medicinal and pharmaceutical products
56	Fertilizers, manufactured
57	Explosives and pyrotechnic products
58	Artificial resins and plastic materials, and cellulose esters etc
59	Chemical materials and products, nes
71	Power generating machinery and equipment
72	Machinery specialized for particular industries
73	Metalworking machinery
74	General industrial machinery and equipment, nes, and parts of, nes
75	Office machines and automatic data processing equipment
77	Electric machinery, apparatus and appliances, nes, and parts, nes
87	Professional, scientific, controlling instruments, apparatus, nes
88	Photographic equipment and supplies, optical goods; watches, etc
95	Armoured fighting vehicles, war firearms, ammunition, parts, nes
E: Human-capital intensive products (*10 sectors*)	
53	Dyeing, tanning and colouring materials
55	Oils and perfume materials; toilet and cleansing preparations
62	Rubber manufactures, nes
64	Paper, paperboard, and articles of pulp, of paper or of paperboard
67	Iron and steel
69	Manufactures of metals, nes
76	Telecommunications, sound recording and reproducing equipment
78	Road vehicles
79	Other transport equipment
89	Miscellaneous manufactured articles, nes
F: Non-classified products (*3 sectors*)	
91	Postal packages not classified according to kind
93	Special transactions, commodity not classified according to class
96	Coin (other than gold coin), not being legal tender

Appendix 3.5: Transition probability matrix for four dropped countries: category I, 1-year

Algeria	No	Weak	Medium	Strong	Libya	No	Weak	Medium	Strong
No	0.998	0.002	0.000	0.000	No	0.993	0.006	0.001	0.000
Weak	0.667	0.333	0.000	0.000	Weak	0.269	0.692	0.038	0.000
Medium	0.333	0.000	0.667	0.000	Medium	0.000	0.111	0.889	0.000
Strong	0.000	0.000	0.000	1.000	Strong	0.000	0.000	0.000	1.000
Mali	No	Weak	Medium	Strong	Mauritius	No	Weak	Medium	Strong
No	0.992	0.004	0.000	0.004	No	0.985	0.015	0.000	0.000
Weak	0.500	0.333	0.000	0.167	Weak	0.054	0.882	0.065	0.000
Medium	0.000	1.000	0.000	0.000	Medium	0.000	0.167	0.778	0.056
Strong	0.029	0.000	0.029	0.943	Strong	0.000	0.000	0.000	1.000

Appendix 3.6: Transition probability matrix for individual African countries: 1-year

Category I: BI-1, 2, 4					Category II: Quartiles				
Angola	No	Weak	Medium	Strong	Angola	1	2	3	4
No	0.997	0.002	0.002	0.000	1	0.817	0.140	0.043	0.000
Weak	0.500	0.500	0.000	0.000	2	0.167	0.736	0.080	0.017
Medium	0.000	0.333	0.333	0.333	3	0.012	0.109	0.818	0.061
Strong	0.000	0.000	0.050	0.950	4	0.000	0.000	0.069	0.931
Theory	0.960	0.006	0.004	0.030	Theory	0.224	0.227	0.261	0.288
Empiric	0.960	0.008	0.004	0.028	Empiric	0.250	0.250	0.250	0.250
Cote d'Ivoire	No	Weak	Medium	Strong	Cote d'Ivoire	1	2	3	4
No	0.987	0.013	0.000	0.000	1	0.913	0.083	0.005	0.000
Weak	0.191	0.660	0.149	0.000	2	0.086	0.847	0.067	0.000
Medium	0.000	0.060	0.930	0.010	3	0.009	0.057	0.891	0.043
Strong	0.000	0.000	0.043	0.957	4	0.000	0.000	0.048	0.952
Theory	0.784	0.053	0.132	0.031	Theory	0.265	0.240	0.262	0.232
Empiric	0.741	0.058	0.119	0.083	Empiric	0.251	0.249	0.251	0.249
Egypt	No	Weak	Medium	Strong	Egypt	1	2	3	4
No	0.962	0.038	0.000	0.000	1	0.907	0.093	0.000	0.000
Weak	0.110	0.733	0.144	0.014	2	0.078	0.836	0.082	0.004
Medium	0.000	0.143	0.773	0.084	3	0.004	0.058	0.837	0.101
Strong	0.000	0.011	0.079	0.910	4	0.000	0.007	0.076	0.916
Theory	0.479	0.167	0.170	0.184	Theory	0.194	0.217	0.262	0.327
Empiric	0.674	0.139	0.105	0.082	Empiric	0.251	0.250	0.250	0.250
Ethiopia	No	Weak	Medium	Strong	Ethiopia	1	2	3	4
No	0.975	0.012	0.009	0.003	1	0.785	0.182	0.033	0.000
Weak	0.333	0.444	0.167	0.056	2	0.085	0.754	0.153	0.008
Medium	0.000	0.094	0.625	0.281	3	0.000	0.099	0.838	0.063
Strong	0.010	0.000	0.067	0.923	4	0.000	0.000	0.048	0.952
Theory	0.504	0.027	0.092	0.377	Theory	0.070	0.177	0.312	0.441
Empiric	0.692	0.036	0.065	0.207	Empiric	0.250	0.250	0.250	0.249
Gabon	No	Weak	Medium	Strong	Gabon	1	2	3	4
No	0.996	0.004	0.000	0.000	1	0.823	0.149	0.021	0.007
Weak	0.091	0.727	0.182	0.000	2	0.137	0.799	0.058	0.007
Medium	0.000	0.083	0.667	0.250	3	0.020	0.108	0.784	0.088
Strong	0.000	0.000	0.065	0.935	4	0.014	0.000	0.076	0.910
Theory	0.670	0.028	0.062	0.240	Theory	0.274	0.305	0.189	0.232
Empiric	0.906	0.020	0.020	0.053	Empiric	0.250	0.250	0.250	0.250
Kenya	No	Weak	Medium	Strong	Kenya	1	2	3	4
No	0.968	0.028	0.004	0.000	1	0.920	0.074	0.007	0.000
Weak	0.141	0.727	0.133	0.000	2	0.059	0.861	0.077	0.003
Medium	0.000	0.142	0.775	0.083	3	0.003	0.065	0.847	0.085
Strong	0.000	0.000	0.054	0.946	4	0.000	0.000	0.069	0.931
Theory	0.543	0.124	0.130	0.203	Theory	0.170	0.215	0.270	0.344
Empiric	0.690	0.111	0.102	0.098	Empiric	0.250	0.250	0.250	0.250

Appendix 3.7: Transition probability matrix for individual African countries: 1-year (continued)

Category I: BI-1, 2, 4					Category II: Quartiles				
Madagascar	No	Weak	Medium	Strong	Madagascar	1	2	3	4
No	0.968	0.028	0.003	0.001	1	0.876	0.124	0.000	0.000
Weak	0.260	0.562	0.178	0.000	2	0.102	0.777	0.117	0.004
Medium	0.013	0.130	0.714	0.143	3	0.004	0.094	0.826	0.075
Strong	0.000	0.000	0.099	0.901	4	0.004	0.004	0.060	0.932
Theory	0.652	0.074	0.109	0.166	Theory	0.207	0.229	0.260	0.303
Empiric	0.745	0.068	0.076	0.112	Empiric	0.250	0.250	0.250	0.249
Malawi	No	Weak	Medium	Strong	Malawi	1	2	3	4
No	0.986	0.009	0.003	0.002	1	0.823	0.167	0.010	0.000
Weak	0.116	0.744	0.140	0.000	2	0.138	0.708	0.154	0.000
Medium	0.000	0.077	0.769	0.154	3	0.000	0.125	0.792	0.083
Strong	0.000	0.029	0.000	0.971	4	0.000	0.000	0.071	0.929
Theory	0.587	0.070	0.050	0.293	Theory	0.170	0.218	0.282	0.330
Empiric	0.824	0.053	0.035	0.087	Empiric	0.250	0.250	0.250	0.250
Morocco	No	Weak	Medium	Strong	Morocco	1	2	3	4
No	0.983	0.015	0.001	0.000	1	0.916	0.084	0.000	0.000
Weak	0.118	0.814	0.069	0.000	2	0.057	0.846	0.096	0.000
Medium	0.014	0.087	0.841	0.058	3	0.000	0.074	0.877	0.049
Strong	0.000	0.000	0.025	0.975	4	0.000	0.000	0.042	0.958
Theory	0.685	0.088	0.068	0.159	Theory	0.151	0.222	0.289	0.338
Empiric	0.744	0.090	0.061	0.106	Empiric	0.251	0.250	0.250	0.250
Mozambique	No	Weak	Medium	Strong	Mozambique	1	2	3	4
No	0.982	0.008	0.008	0.002	1	0.877	0.116	0.000	0.006
Weak	0.333	0.542	0.083	0.042	2	0.173	0.698	0.123	0.006
Medium	0.031	0.219	0.594	0.156	3	0.012	0.167	0.753	0.068
Strong	0.000	0.000	0.077	0.923	4	0.000	0.000	0.099	0.901
Theory	0.775	0.037	0.049	0.139	Theory	0.376	0.253	0.197	0.174
Empiric	0.790	0.039	0.050	0.120	Empiric	0.251	0.249	0.251	0.249
Senegal	No	Weak	Medium	Strong	Senegal	1	2	3	4
No	0.969	0.027	0.003	0.001	1	0.860	0.132	0.007	0.000
Weak	0.200	0.700	0.075	0.025	2	0.086	0.792	0.115	0.007
Medium	0.000	0.098	0.738	0.164	3	0.007	0.081	0.849	0.063
Strong	0.006	0.000	0.076	0.917	4	0.000	0.004	0.048	0.948
Theory	0.598	0.084	0.095	0.222	Theory	0.141	0.206	0.283	0.369
Empiric	0.722	0.075	0.059	0.144	Empiric	0.251	0.250	0.250	0.250
South Africa	No	Weak	Medium	Strong	South Africa	1	2	3	4
No	0.975	0.025	0.000	0.000	1	0.913	0.075	0.011	0.000
Weak	0.102	0.828	0.070	0.000	2	0.067	0.855	0.078	0.000
Medium	0.008	0.053	0.878	0.061	3	0.008	0.042	0.905	0.046
Strong	0.000	0.000	0.083	0.917	4	0.000	0.000	0.037	0.963
Theory	0.592	0.136	0.157	0.115	Theory	0.157	0.169	0.303	0.371
Empiric	0.668	0.121	0.123	0.088	Empiric	0.251	0.250	0.250	0.250

Appendix 3.8: Transition probability matrix for individual African countries: 1-year (continued)

Category I: BI-1, 2, 4					Category II: Quartiles				
Togo	No	Weak	Medium	Strong	Togo	1	2	3	4
No	0.968	0.030	0.002	0.000	1	0.854	0.139	0.007	0.000
Weak	0.138	0.638	0.224	0.000	2	0.168	0.720	0.105	0.007
Medium	0.043	0.196	0.543	0.217	3	0.000	0.109	0.776	0.116
Strong	0.000	0.016	0.095	0.889	4	0.000	0.007	0.085	0.908
Theory	0.601	0.111	0.097	0.190	Theory	0.254	0.221	0.225	0.299
Empiric	0.708	0.102	0.077	0.113	Empiric	0.250	0.250	0.250	0.250
Tunisia	No	Weak	Medium	Strong	Tunisia	1	2	3	4
No	0.985	0.015	0.000	0.000	1	0.923	0.077	0.000	0.000
Weak	0.085	0.884	0.031	0.000	2	0.064	0.846	0.086	0.004
Medium	0.000	0.132	0.789	0.079	3	0.000	0.054	0.903	0.043
Strong	0.000	0.000	0.020	0.980	4	0.000	0.000	0.029	0.971
Theory	0.718	0.130	0.031	0.121	Theory	0.138	0.164	0.272	0.426
Empiric	0.761	0.116	0.033	0.090	Empiric	0.250	0.250	0.250	0.250
Uganda	No	Weak	Medium	Strong	Uganda	1	2	3	4
No	0.968	0.028	0.003	0.000	1	0.833	0.163	0.005	0.000
Weak	0.121	0.672	0.207	0.000	2	0.097	0.739	0.164	0.000
Medium	0.019	0.135	0.692	0.154	3	0.000	0.075	0.851	0.075
Strong	0.009	0.009	0.043	0.940	4	0.000	0.000	0.044	0.956
Theory	0.511	0.097	0.110	0.282	Theory	0.076	0.132	0.295	0.497
Empiric	0.719	0.075	0.067	0.139	Empiric	0.251	0.249	0.251	0.249
Zambia	No	Weak	Medium	Strong	Zambia	1	2	3	4
No	0.968	0.024	0.005	0.003	1	0.827	0.149	0.019	0.005
Weak	0.262	0.554	0.154	0.031	2	0.141	0.743	0.117	0.000
Medium	0.074	0.222	0.519	0.185	3	0.015	0.097	0.791	0.097
Strong	0.000	0.000	0.131	0.869	4	0.005	0.005	0.088	0.903
Theory	0.740	0.071	0.064	0.125	Theory	0.220	0.234	0.268	0.278
Empiric	0.766	0.076	0.059	0.099	Empiric	0.251	0.250	0.250	0.250

Appendix 3.9: Mobility indices for individual countries

	Category I: BI-1, 2, 4				Category II: Quartiles		
Country	M1	M2	M3	M4	M1	M2	M3
Angola	0.407	0.492	0.850	0.025	0.233	0.188	0.568
Cote d'Ivoire	0.155	0.180	0.431	0.026	0.133	0.104	0.355
Egypt	0.207	0.145	0.529	0.045	0.168	0.130	0.433
Ethiopia	0.344	0.581	0.774	0.034	0.224	0.174	0.553
Gabon	0.225	0.139	0.574	0.008	0.228	0.200	0.555
Kenya	0.195	0.140	0.507	0.032	0.147	0.115	0.388
Madagascar	0.285	0.253	0.682	0.042	0.196	0.148	0.497
Malawi	0.177	0.132	0.463	0.024	0.250	0.184	0.607
Morocco	0.129	0.125	0.352	0.016	0.134	0.100	0.361
Mozambique	0.320	0.358	0.732	0.044	0.257	0.206	0.619
Senegal	0.225	0.205	0.559	0.033	0.184	0.138	0.470
South Africa	0.134	0.111	0.360	0.029	0.121	0.091	0.328
Togo	0.321	0.236	0.753	0.051	0.247	0.185	0.598
Tunisia	0.121	0.091	0.333	0.012	0.119	0.083	0.324
Uganda	0.242	0.183	0.608	0.040	0.207	0.140	0.524
Zambia	0.363	0.371	0.802	0.069	0.245	0.196	0.589

Chapter 4 Regions, agglomerations, and cities in China[54]

4.1 Introduction

China has seen a dramatic transition from a central planned economy to a market-oriented economy since 1978. During the economic reform, labor force has become increasingly mobile across regions, which formed two main strands of migrants. First is the rural-urban migration. In the beginning, the agricultural reforms in rural areas and more open markets for food and housing in cities made it possible for farmers to enter cities (Zhu and Luo, 2010). The development of private enterprises enabled rural-urban migrants to seek jobs and earn their living in cities (Zhu and Luo, 2010). With the development of urban economies and the rise in rural-urban income inequalities, an increasing number of workers had left agricultural production in rural areas and moved to manufacturing sectors in urban areas (Zhang and Song, 2003; Du, Park and Wang, 2005; Chen, Jin and Yue, 2010). Second is the migration among different urban areas. Benefiting from preferential policies and enormous FDI inflows created by economic reforms, certain coastal regions and cities, such as Shenzhen and Shanghai, have been growing rapidly. Driven by these policies and investments, resource reallocations and enlarged income gaps among different urban regions enticed a considerable proportion of urban labors, especially high-skill workers, to migrate from underdeveloped cities to more prosperous cities with a large concentration of human capital and technological changes (Fu and Gabriel, 2012).

Economic reforms also contribute to the agglomeration of industries. Explaining the spatial concentration of Chinese manufacturing from 1980 to 1995, Wen (2004) finds that (consistent with the new geographic theory) Chinese manufacturing is highly geographically concentrated in several coastal regions. Likewise, Ge (2009) finds that export-oriented and foreign-investment sectors have a higher degree of agglomeration than other sectors in the period 1985 to 2005. The market-oriented distribution of mobile laborers and industries seems to be related. On the one hand, the agglomeration of

[54] This chapter is based on joint research with Steven Brakman and Charles van Marrewijk.

industries attracts a large number of mobile laborers. On the other hand, both the rural-urban and urban-urban migrant laborers contribute to the rapid urbanization and agglomeration of economic activities, which in turn accelerates economic growth (Young, 2003).[55]

Previous studies investigate the distribution of migrant workers and industries based on the regional and agglomeration theoretical framework. In addition, most research focuses on the early stages of the economic reform in the 1980s and 1990s. In contrast, this paper is based on urban economics and investigates whether larger cities are relatively skill abundant and specialize more in skill-intensive sectors and occupations as time goes by. This is not only of academic and policy relevance but also provides a good setting to study the evolution of the economy within cities.

Based on the theoretical framework developed by Davis and Dingel (2013) and using data from the Chinese census of population in 2000 and 2010, we empoly an elasticity test and a pairwise comparison test to identify the interactive relationships between city size and skills, sectors, and occupations of Chinese cities in 2000 and 2010. The results of both tests show that larger cities are relatively more skill abundant in both 2000 and 2010. The results for sectors and occupations confirm this only in 2010, however. We take this as an indication that the Chinese economy is becoming more market-oriented over time. These dynamic comparison is absent in the study of Davis and Dingel (2013). Our paper also contributes to the literature of urbanization in less developed countries.

The remainder of this paper is organized as follows. Section 4.2 reviews the related studies and the theoretical framework. Section 4.3 sets out the methodology of the elasticity test and the pairwise comparison test. Section 4.4 discusses data sources. Section 4.5 presents the results of identifying the relationships between city size and skills, sectors, and occupations. Section 4.6 offers concluding remarks.

---

[55] Au and Henderson (2006) and Whalley and Zhang (2007) claim that there are some restrictions of laborers' migration impeding the agglomeration degree of industries and economic growth in China. However, the restrictions were actually limited. Because, first of all, the return migration is of limited scale for rural-urban migrant laborers and the average returnee is older (Zhao, 2002). The older migrations are not workers anymore, so their living locations do not affect the agglomeration. Secondly, the migration restrictions were low for workers with high skills in urban-urban migration (Liu, 2005).

4.2 Theoretical framework

4.2.1 Related studies

This paper is related to two strands of prior literature. One strand of literature focuses on the distribution of skills across cities. Examining the distribution of skills across cities, Glaeser (1999), Mori and Turrini (2005), Glaeser and Resseger (2010) find that workers of higher skills are inclined to stay in larger cities. However, all these studies measure skills by educational levels. Bacolod, Blum, and Strange (2009) made a seminal contribution by grouping worker skills into three categories for US workers: cognitive, people, and motor skills. By comparing the distribution of skills across four groups of cities (small, medium, large, and very large cities), they find that, compared to the smaller cities, larger cities are more skill abundant but to a modest degree. The difference in the skill categories of cognitive, people and motor skills across cities is smaller compared to differences in educational levels. Though the category approach of skills is new and innovative to the literature of urban economics, this approach can hardly be applied to other countries because of data availability.

Another strand of literature is about the sectoral distribution across cities. Henderson (1974) provides a classic general equilibrium model to analyze the relationship between the optimal city size and city sector composition. He argues that the optimal city size is characterized by the trade-off between the benefits and costs of laborers. However, the trade-off varies with the different types of specialized production in the city due to different degrees of economies of scales that support different levels of commuting and congestion costs. Henderson (1983) further empirically explores the relationship between the industrial bases and the city size. Using the data of the United States in 1970 and the “back-of-the-envelope” method, he investigates how the employment of one industry varies with city size. Regressing the share of employment in each industrial category on the urban population in cities, he finds that some manufacturing activities, business services, and the sectors of finance, insurance and real estate appear to concentrate in larger cities. The only exception is the resource-based manufacturing which tends to decline with city size. Henderson (1997) extends the empirical work to other economies, such as Brazil, Japan and Korea. By utilizing the share of local employment or productivity of some specific sectors against the cities’ population, he finds similar production patterns of the medium-size cities versus larger-size cities in all of these countries. To be specific, the medium-size cities tend to be relatively more specialized in manufacturing activities, especially in the low-skill intensive industries, while the larger cities tend to contain the high-tech and diversified manufactures, business services, and R&D activities. He attributes the reason for the second pattern to the large demand for local diverse labor and product markets in these economic activities. Moreover, Holmes and Stevens (2004) empirically examine the spatial distribution of economic activities in North America. By analyzing the distribution of the major sectors in each of the four urbanization quartiles they find substantial variations in the pattern of specialization based on market size proxied by the concentration of population.[56] In particular, agriculture, mining, manufacturing, and utilities concentrate in smaller cities. In contrast, transportation, wholesale trade, real estate, finance & insurance, management, and professional services trend to concentrate in larger cities (consistent with previous studies, see Henderson, 1983 and 1997).

Combining these studies, some researchers seek to jointly analyze the distributions of skills and sectors across cities. Abdel-Rahman and Anas (2004) use two homogenous skill levels, namely unskilled and skilled labor. In contrast, Behrens, Duranton and Robert-Nicoud (2013) develop a theoretical model linking labor skills, occupation choice, and city size assuming a continuum of skills. Both studies assume that all individuals are indifferent within a city in terms of their location. To overcome this weakness Davis and Dingel (2013) develop a model of a system-of-cities to explore the joint relationship between the skills distribution across cities and the sectoral employment distribution across cities. In contrast to previous system-of-cities theories based on polarized sectoral composition, that is specialized and perfectly diversified cities (Henderson, 1974; Abdel-Rahman and Anas, 2004; Helsley and Strange, 2012), the authors develop a case in which cities are incompletely specialized.

The spatial distribution of industries has been explained by various theories in terms of agglomeration (Marshall, 1920; Krugman, 1991; Holmes, 1998 and Arzaghi and Henderson, 2008). Recent empirical studies address the significance of agglomeration, localization, and coagglomeration of industries, which are relative to the locations of

[56] Urbanization quartiles refer to four groups of cities ranked by population, where each of the group contains 25 percent of total U.S. or Canadian employment.

industries being uniformly allocated to the proportion of local population (Ellison and Glaeser, 1997; Duranton and Overman, 2005; Ellison, Glaeser and Kerr, 2010). Obtaining similar results as previous empirical studies, Davis and Dingel (2013), however, rely on urbanization economies and individuals' comparative advantage. They predict that larger cities will be more skill abundant and specialize relatively more in skill-intensive activities, because larger cities own more desirable locations that attract more high-skill laborers, who normally work in more skill-intensive sectors.

4.2.2 Model structure

Our empirical work is based on the theoretical model of Davis and Dingel (2013). They develop a fairly general framework in which L heterogeneous individuals with a continuum of skills s sort over a continuum of (intermediate good) sectors σ by choosing from a continuum of locations δ within C discrete cities, $c \in \mathbb{C} = \{1, .., C\}$. Their objective is to maximize utility U, which is equal to disposable income, given by the difference between the individual's value of productivity $q(\mathrm{c}, \delta, \sigma; \mathrm{s})p(\sigma)$, where q is productivity and p is the price of the intermediate good, and the rental rate $\mathrm{r}(\mathrm{c}, \delta)$, see equation (1). The rental rate only depends on the city and the location within the city. An individual's productivity depends on the city-level total factor productivity $A(c)$, which is taken as given by the individuals but depends on the city's size and the distribution of skills within the city, interacted with location $\mathrm{D}(\delta)$ and the choice of sector combined with skills $H(\mathrm{s}, \sigma)$, multiplicatively.

$$U(s, c, \delta, \sigma) = q(.)p(.) - r(.) = A(c)D(\delta)H(s, \sigma)p(\sigma) - r(c, \delta) \qquad (1)$$

As a normalization, higher δ locations in a city are less attractive/productive, so $D'(\delta) < 0$. One can think of commuting costs to the central business district, but an alternative interpretation of the model is the desirability of a location because of its consumption value. The function H is assumed to be strictly log-supermodular (in s and σ) and strictly increasing in skills.[57] This ensures that higher skilled individuals are more productive and also *relatively* more productive in higher σ (more skill-intensive) sectors. Individuals

[57] That is: $s > s', \sigma > \sigma' \Longrightarrow H(s, \sigma)H(s', \sigma') > H(s, \sigma')H(s', \sigma)$.

supply one unit of labor inelastically and pay rent to absentee landlords, who engage in Bertrand competition.

In a competitive equilibrium individuals choose location within the city and the sector to work in independently as these enter the objective function separably. We order the system of cities in terms of total factor productivity such that $A(C) \geq A(C-1) \geq .. \geq A(1)$. Define the *attractiveness* γ of a location δ within a city c as: $\gamma = A(c)D(\delta)$. If Sarah, with skill level s, is indifferent between living in Shanghai (city c, say, with population 23 million) and Suzhou (city c', say, with population 4.1 million) it must mean that the attractiveness of the two locations is the same: $A(c)D(\delta) = A(c')D(\delta')$. Since the bigger city (Shanghai) has the higher total factor productivity ($A(c) > A(c')$), this means that Sarah can choose between a not-so-good location in Shanghai and a wonderful location in Suzhou ($D(\delta) < D(\delta')$). Since the people with the highest skill levels can afford to choose the most attractive locations, there will be a range of high-skilled people living in Shanghai that cannot be found in Suzhou, followed by a range of people with similar skill levels found in both cities. Since higher-skilled people work in the more skill-intensive sectors, larger cities contain relatively more skill-intensive sectors.

Davis and Dingel (2013) show that, under a regularity condition (namely that the supply of locations in a city is decreasing and log-concave), the distribution of skills over cities (say $f(s,c)$, which is integrated over sectors and locations within the city) is log-supermodular, see equation (2). Moreover, the same holds for output, employment, and revenue from a sector perspective. The inequality in equation (2) satisfies the monotone likelihood ratio property, which means that the relative returns to increasing skills (s) or the skill-intensity of sectors (σ) are increasing in city size (Milgrom, 1981; Costinot, 2009). This allows us to evaluate the main implications of the model using two simple empirical tests, as discussed in the next section.

$$f(s,c)f(s',c') \geq f(s,c')f(s',c), \text{ for } c \geq c' \text{ and } s \geq s' \tag{2}$$

In deriving the above relationship we imposed a competitive equilibrium in which laborers are allowed to move freely. Since China has been engaged in a long transformation process going from a centrally-planned economy to a more market-oriented economy ever since

Deng Xiaoping started the Economic Reform process in 1978, we expect the predictive power to improve as time progresses. Moreover, since there were many restrictions on labor mobility in the past (the Hukou system) which have only gradually been lifted (while some restrictions are still in place to this day), we again expect that the predictive power of the model improves as time progresses. Indeed, in the discussion below, we will interpret changes over time regarding the predictive power of the model as an indication of China's move to a more market-oriented economy characterized by more labor mobility. To summarize the discussion, we have the following

Hypotheses In a competitive equilibrium with mobile workers
H1*:* Larger cities are relatively more skill abundant.
H2*:* Larger cities produce relatively more in skill-intensive sectors.
H3*:* The empirical tests of the model improve over time.

4.3 Empirical methodology

To identify the effect of the city size on the distribution of skilled laborers and skill-intensive sectors, we use two simple empirical tests, namely the "elasticity test" and the "pairwise comparisons test", as discussed below.

4.3.1 Elasticity test

Hypothesis 1 and 2 state that larger cities are relatively skill abundant and produce relatively more in skill-intensive sectors. In other words, the city-population elasticity of the skill type should be increasing in skill levels. Similarly, the city-population elasticity of sectoral employment should be increasing in the skill intensity of sectors. In our empirical work we order the skill-intensity either by sector σ or by occupation o and use the following regression:

$$lnf(v,c) = \beta_{v0} + \beta_{v1}\alpha_v + \beta_{v2}lnL(c) + \beta_{v3}\,\alpha_v * lnL(c) + \epsilon_{v,c}, \quad where\ v = s,\sigma,o \quad (3)$$

Where s, σ, and o denotes the skill level, sector, or occupation, $lnf(v,c)$ is the natural logarithm of the distribution (of skills, sectors, or occupation) across cities, α_v represents the fixed effect, $lnL(c)$ is the natural logarithm of the city population, and the β's are estimated coefficients. To estimate the elasticities, we interact fixed effects with city population, allowing the impact of city size to depend on different groups of skills, sectors, and occupations. In view of the derived log-supermodularity of the density functions in the model, the crucial estimated coefficient β_{v3} on the city-population elasticity of the skills, sectors and occupations can be ordered as follows: $\beta_{v3} \geq \beta_{v'3} \leftrightarrow v \geq v'$. This regression can be understood as a first-order Taylor approximation.

4.3.2 Pairwise comparison test

Suppose we have empirical information on the distribution of 4 types of skills, ranked according to skill level, across 40 cities, ranked according to size. We can then directly compare for any two arbitrary cities and two skill types whether or not inequality (2) holds, that is verify if the larger city has relatively more workers of the higher skill type. Say we call the comparison a 'succes' if the condition holds (value = 1) and a 'failure' if not (value = 0). We can compare 40 cities in $(40\times39)/2 = 780$ different ways and 4 skill types in $(4\times3)/2 = 6$ different ways for a total of $780\times6 = 4680$ comparisons. The extent to which the average succes rate exceeds the random distribution benchmark of 0.5 can then be taken as an indication regarding the sorting-predictive power of the model. Similarly, if we have various types of sectors or occupations ranked according to skill level.

We can, however, go a bit further. First, in view of the general prediction of the model and the usual problems encountered when we are collecting and processing data, we should expect the inequality comparison between a very large city (such as Shanghai with 23 million people) and a much smaller city (such as Wuhai in Inner Mongolia with 0.5 million people) to hold almost surely. In contrast, when we compare two similar-sized cities, such as Wuhai (532,902 persons) with Nujiang in Yunnan (534,337 persons), we are not so sure. We will therefore reported 'weighted' succes rates, where we use the difference in log population for the two compared cities as weight. Second, we do not have to restrict ourselves to comparing individual cities. We can also compare groups of cities in 'bins' of different size. Suppose we have two distinct sets of cities C and C' with the smallest city in C being bigger than the biggest city in C' and two skill types with $v > v'$. Inequality (2) then also holds for the bin:

$$\sum_{c\in C} f(v,c) \sum_{c'\in C'} f(v',c') \geq \sum_{c\in C} f(v',c) \sum_{c'\in C'} f(v,c') \tag{4}$$

This inequality implies that if the cities are grouped into a series of bins ordered by city size, then in any pairwise comparison of two bins and two skills the bin containing the larger cities has relatively more of the high-skilled workers. Similarly for sectors and occupations. When we create 2 bins we have just 1 comparison (large versus small cities). When we create 4 bins we have 6 comparisons, and so on. In the analysis below we divide the cities into 2, 4, 10, 30, 90, and individual bins.[58] If m is the number of bins and n is the number of skills (sectors or occupations) the total number of pairwise comparisons is thus $\frac{m(m-1)}{2} \times \frac{n(n-1)}{2}$. We report both the unweighted and weighted succes rate of the pairwise comparisons per bin.[59]

4.4 Data

4.4.1 The administrative division of locations

Our primary data sources are the population census of 2000 and the population census of 2010. The administrative division of Mainland China consists of five levels, but our dataset only covers the top three levels: the provincial level, the prefectural level, and the county level.[60] There are different types of county levels, such as 'district' and 'county' proper, where district is urban-based while county is rural-based. We therefore define two 'city' levels and one 'regional' level below to analyze the sorting of skills, sectors, and occupations over different locations. We label these *Regions*, *Agglomerations*, and *Cities*, see Table 4.1.

First, we identify *Regions* at the prefectural level, which include all seven types of county-level administrative divisions (listed in Table 4.1 from *District* to *Adm. committee*).[61] In terms of coverage, *Region* accounts for more than 98 percent of the total population in both 2000 and 2010.[62] There were 338 regions in 2000 (334 prefectural levels and 4 municipalities) and 337 regions in 2010 (333 prefectural levels and 4 municipalities). Note

[58] Individual bins consist of one city per bin.

[59] We use the difference of the log of the average population in a bin as weight. `

[60] Levels 4 and 5 are the township level and the village level.

[61] There are four municipalities in China at the provincial level (Shanghai, Beijing, Tianjin, and Chongqing) which are comparable to the prefectural levels in other provinces. These four are also classified as *Region*.

[62] Some county-level divisions are administrated by their provinces directly. In that case, the information of the divisions is excluded from the statistic of the prefectural levels. The population share of these county-level divisions is about 1.7 percent, which explains why coverage is not 100 percent of total population.

that *Region* covers both urban and rural areas. It is thus questionable to apply the sorting model discussed in section 4.2 to the regional level because this assumes that mobile workers are choosing between different regions (rather than cities) and the whole region (including various cities and rural areas) is viewed as one location decision. We nonetheless report results at the regional level below in order to be able to compare with the agglomeration and city level and to discuss changes over time.

Table 4.1: Summary of Chinese administrative division at prefectural and county level

Type			Administrative division	2000			2010		
				Population share (%)	Cum.	Num.	Population share (%)	Cum.	Num.
1	2	3	Prefectural level + Municipalities	98.3		338	98.3		337
		City	District	26.3	26.3	803	34.7	37.7	861
	Agglomeration		County-level city	20.3	46.3	389	17.3	52.0	353
Region			County	48.4		1489	43.3		1460
			Auto. county*	2.4		109	2.2		110
			Banner	0.9		52	0.8		52
			Special district	0.0		1	0.0		1
			Adm. committee*	-	98.3	-	0.0	98.3	3

Source: Chinese census of population 2000, 2010.
Note: The short names with star are Autonomous county and Administrative committee, respectively.

Second, we identify an equal number of *Agglomerations* at the prefecture level. This is a subset of *Region* excluding all 'rural' type counties. In particular, we only include *district* and *county-level city*. The share of the total population living in Agglomerations rose from about 46 percent in 2000 to 52 percent in 2010, partially because of direct migration decisions and partially because of changes in administrative division (as a consequence of migration). By construction, Agglomeration is a cluster of urban areas that is viewed to operate as a consistent whole. Since it is a more coherent location definition, the model discussed in section 4.2 should be more or less applicable at the agglomeration level.

Third, we identify an equal number of *Cities* at the prefecture level. This is a subset of *Agglomeration* consisting only of *districts*. This more narrowly defined location thus excludes the *county-level cities*, which could be viewed as more or less independent satellites rather than a true part of the location itself. The share of the total population living in *Cities* rose from about 26 percent in 2000 to 38 percent in 2010, again partially because of direct migration decisions and partially because of changes in administrative division (as a consequence of migration). Since *Cities* is the most coherent location definition, the model discussed in section 4.2 should be most applicable at the City level.

Figure 4.1: Yancheng prefecture; Jiangsu province, China, 2010

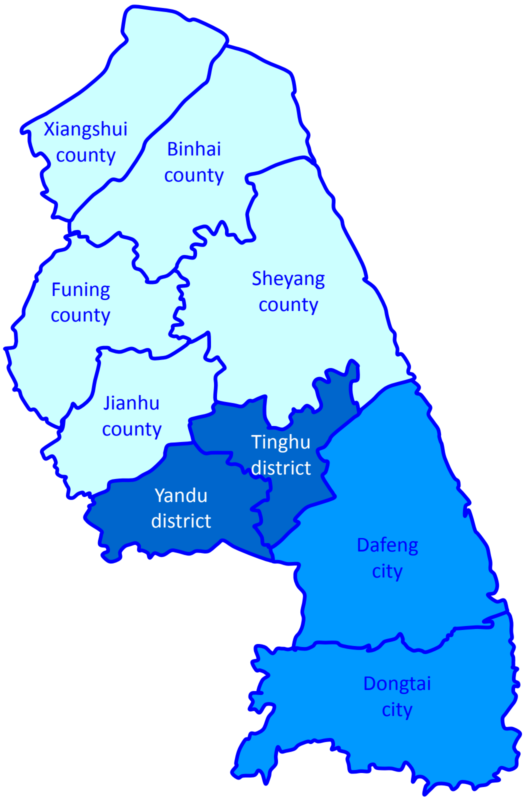


Figure 4.1 illustrates the various location definitions for Yancheng prefecture in the east-coastal province of Jiangsu (close to Shanghai) in 2010. The area of Yancheng prefecture is almost 17,000 km^2, roughly the size of Swaziland or half the size of the Netherlands. Yancheng prefecture consists of 9 county-level sub-regions, namely 2 districts, 2 county-level cities, and 5 (rural) counties. Yancheng *Region* consists of the population of all 9 counties, or about 7.3 million people in total. Yancheng *Agglomeration* consists of the two districts (Yandu and Tinghu) and the two county-level cities (Dafeng and Dongtai), or about 3.3 million people (46 percent of the total population). Yancheng *City* only consists of the two districts Yandu and Tinghu, or about 1.6 million people (22 percent of the total population). The definition thus becomes more concentrated and more coherent as we go from region to agglomeration to city.

4.4.2 Skills

As is common in the literature, we use educational attainment as a proxy for skills. The Chinese census of population (2000 and 2010) categorizes six groups of educational attainments, related to the number of years of schooling. We aggregate the county-level educational data into the three types of locations and calculate the population share of each educational group in the total population of China (see Table 4.2).[63] Two observations are clear upon inspecting this table across time and location type.

First, a comparison across time shows that the education level is *rising over time*: the population share is falling for the two lowest education levels and rising for the three highest education levels for all location types.[64] At the *Region* level the population share of illiterates falls, for example, from 8 percent to less than 5 percent and the population share of at least Bachelor rises from 1.4 to 4 percent.

Second, a comparison across location type shows that the education level is highest in *Cities* and lowest in *Regions*, with *Agglomerations* in between: the population share is falling for the three lowest education levels and rising for the three highest education levels as we move from Regions to Agglomerations to Cities in both time periods. In 2010, for example, the population share with Primary school falls from 28.7 percent at the Region level to 23.4 percent at the Agglomeration level to 20.2 percent at the City level. Similarly, the population share for College rises from 5.5 at the Region level to 7.6 at the Agglomeration level to 9.4 at the City level.

[63] In 2000, there were two additional educational groups, literacy class and technical school. We do not include them in this table since they are excluded in 2010. The data on educational attainment only includes the population of at least six years old.

[64] The third education level is rising for Regions, stable for Agglomeration, and falling for Cities.

Table 4.2: Population shares of skill group by educational attainment in 2000 and 2010(%)

Education	years	Region 2000	Region 2010	Agglomeration 2000	Agglomeration 2010	City 2000	City 2010
Illiterate	0	8.0	4.9	6.3	3.6	5.8	3.2
Primary school	6	40.3	28.7	34.5	23.4	28.9	20.2
Middle school	9	38.7	41.8	40.4	40.6	40.1	38.1
High school	12	9.0	15.1	12.0	18.4	14.8	20.4
College	15	2.6	5.5	4.2	7.6	6.2	9.4
Bachelor	16+	1.4	4.0	2.6	6.5	4.3	8.7
Total % of spatial unit		100	100	100	100	100	100
As % of total population		84.1	88.7	39.9	47.5	22.3	31.9

Source: Chinese census of population 2000, 2010; *years* = number of years of schooling.

4.4.3 Sectors and occupations

The distributions of sectors and occupations varies substantially across Chinese locations. In order to examine the interaction with population size we use data on the sectoral and occupational employment from the Chinese census of population (2000 and 2010).[65] The sectors were classified into 15 categories in 2000 and expanded into 20 categories in 2010, while the number of occupations consists of 7 categories in both years.[66] To test the model we order sectors and occupations with respect to the corresponding skill intensities, which we collect from the China Labor Statistical Yearbook (2010). This lists sectoral and occupational employment as proportions of six educational attainments, measured by years of schooling.[67] The breakdown is provided both for the economy as a whole and for urban employment.[68]

[65] The population of a location consists of both registered residents and non-registered residents living there continuously for at least five years.
[66] We drop the sector *International organizations* because it has almost zero employment in 2010.
[67] There is no educational information about sectors and occupations in 2000. Therefore, we order the skill intensity of sectors and occupations only based on the information available in 2010.
[68] Labeled 'Total' and 'Urban', respectively, in the left-hand panel of Table 4.3, see below.

Table 4.3: Average education of employment and population share in each sector

Sector	Average education						Share of working population spatial unit (%)					
	Total			Urban			Region		Agglomeration		City	
	Years	Order 2000	Order 2010	Years	Order 2000	Order 2010	2000	2010	2000	2010	2000	2010
Farming	7.38	1	1	7.73	1	1	64.5	41.4	46.7	25.7	32.5	17.3
Construction	9.03	2	2	9.54	2	2	2.7	4.8	4.0	5.2	4.7	5.3
Public Services	9.44	3	3	9.64	3	3	2.1	1.7	3.5	2.1	5.1	2.3
Mining	9.53	4	4	10.20	5	6	1.0	1.0	1.4	1.1	1.6	1.1
Hotel	9.55	-	5	9.76	-	4	-	2.4	-	2.9	-	3.4
Manufacturing	9.69	5	6	10.14	4	5	12.8	15.0	20.0	19.1	24.0	19.4
Trade	9.95	6	7	10.21	6	7	6.7	8.1	10.0	10.6	13.2	12.2
Transport	10.02	7	8	10.35	7	8	2.6	3.0	3.7	3.8	4.7	4.4
Public Utility	10.66	8	9	10.98	8	9	0.1	0.3	0.2	0.5	0.2	0.6
Real Estate	11.48	9	10	11.63	9	10	0.2	0.6	0.5	1.0	0.8	1.3
Utilities	11.72	10	11	12.06	10	11	0.6	0.6	0.9	0.8	1.2	0.9
Culture[69]	11.85	12	12	12.08	11	12	2.5	0.4	3.2	0.6	4.1	0.8
Business Serv.	12.04	-	13	12.30	-	13	-	14.1	-	18.1	-	20.7
Research	12.91	11	14	13.36	13	16	0.2	0.3	0.4	0.5	0.8	0.6
Computer	12.96	-	15	13.29	-	14	-	0.5	-	0.8	-	1.1
Public Health	13.00	13	16	13.35	12	15	1.1	1.0	1.5	1.3	2.0	1.5
Public Adm.	13.36	14	17	13.60	14	17	2.3	2.2	3.1	2.6	3.8	3.1
Banking	13.64	15	18	13.76	15	18	0.6	0.7	0.9	1.1	1.3	1.4
Education	14.09	-	19	14.36	-	19	-	2.0	-	2.3	-	2.7
As % of identified working population spatial unit							100	100	100	100	100	100
As % of total population							51.2	59.1	23.4	31.6	12.5	21.0

Sources: China Labor statistical yearbook (2010) and Chinese census of population (2010); years = the number of years of schooling; Serv. = Services; Adm. = Administration.

The skill intensity can be calculated by the weighted average years of schooling in each sector and occupation, ordered from low to high in in *Total* and *Urban* levels (see Table 4.3 and Table 4.4, left part).[70] *Total* denotes the skill intensity of total employment, while *Urban* focuses on the employment in urban areas, which includes all districts in prefectural levels and the center of towns below county levels. Generally, the average years of education in urban areas are higher than that of the total areas. Most orders are identical in both *Total* and *Urban* levels with some exceptions. In the following empirical tests, we use the *Total* order in *Region* estimations and *Urban* order in *Agglomeration* and *City* estimations.

---

[69] The sector *Culture* is a joint sector with *Education* in 2000. We use the average years of schooling of *Culture* and *Education* as the skill intensity of *Culture* in 2000, which are 12.97 years and 13.22 years for *Total* and *Urban* areas, respectively (see Table 4.3). The order of *Culture* in 2000 is based on this calculation.

[70] Years of schooling=$\sum_{e=1}^{6} s_e * p_{ie}$, where e is the educational attainment. i denotes the sector or occupation. s_e denotes the years of schooling of each educational attainment. p_{ie} denotes the share of the educational attainment e in the sector or occupation i.

Table 4.4: Average education of employment and population share in each occupation

Occupation	Average education				Share of working population spatial unit (%)					
	Total		Urban		Region		Agglomeration		City	
	Years	Order	Years	Order	2000	2010	2000	2010	2000	2010
Agriculture	7.38	1	7.74	1	64.4	47.8	46.7	31.1	32.4	21.5
Production	9.29	2	9.68	2	16.0	22.9	24.0	28.3	28.0	29.3
Others	9.73	3	10.20	4	0.1	0.1	0.1	0.1	0.1	0.1
Business Serv.	9.82	4	10.07	3	9.2	16.3	13.8	21.9	18.0	25.8
Unit Head	11.72	5	12.12	5	1.7	1.8	2.5	2.6	3.4	3.2
Clerk	12.64	6	12.90	6	3.1	4.3	4.9	6.4	7.2	8.2
Technical Pers.	13.10	7	13.48	7	5.6	6.8	8.0	9.6	10.9	12.0
As % of identified working population spatial unit					100	100	100	100	100	100
As % of total population					51.3	51.1	23.5	26.2	12.6	16.9

Sources: China Labor statistical yearbook (2010) and Chinese census of population (2010); *years* = the number of years of schooling; Serv. = Services; Pers. = Personnel.

The right parts of Table 4.3 and Table 4.4 show the share of each sector and occupation in the total population of China for the three types of locations. For sectors (Table 4.3), *Farming* absorbed the largest share of population (except for *Cities* in 2010), followed by *Manufacturing* in both 2000 and 2010. Although it is hard to compare developments over time because of the identification of 4 new sectors, it is clear that the *Farming* employment fell drastically over time, namely from 65 to 41 percent at the *Region* level, from 47 to 26 percent at the *Agglomeration* level, and from 33 to 17 percent at the *City* level. A comparison across location types is simple for both periods: the working population share in *Farming* falls as we move from *Regions* to *Agglomerations* to *Cities*, while the working population share for all other sectors either rises or is stable.

For occupations (Table 4.4) the changes are straightforward (as there are no occupations added). The largest employment is in the occupation *Agriculture* (again, as with *Farming* for sectors, with the exception of Cities in 2010). The employment in *Agriculture* falls over time, while the employment in all other occupations rises over time for all location types (with the exception of *Unit Heads* in *Cities*). When we compare across location types, employment is falling for *Agriculture* and rising for all other occupations as we move from *Regions* to *Agglomerations* to *Cities* in both periods (except for the 'Others' occupation, which is stable).

4.5 Empirical results

In this section, we utilize two empirical methods to test our hypotheses, namely, whether larger cities are relatively more skill abundant and specialized in skill-intensive sectors and occupations. First, we examine the relationship between city size and the distribution of skills. We find that results strongly confirm the prediction of hypothesis 1 for all three location levels in both 2000 and 2010. We also find that the 2010 results are stronger than the 2000 results. Second, after investigating the distributions of skills, we examine the relationship between the city size and the distribution of sectors and occupations. We find clear evidence that China's sectoral and occupational distribution across cities changed from 2000 to 2010. More specifically, larger cities produced relatively more in higher skill-intensive sectors and occupations only in 2010. We do not find support for this prediction in 2000.

4.5.1 Larger cities are relatively more skilled

A. Elasticity test

This subsection examines the links between city size and the distribution of skills. Table 4.5 reports the results of population elasticities of educational groups for our three types of location. In general, the estimated elasticities confirm that larger locations have relatively more skilled inhabitants: the elasticities are higher for more skilled educational groups at the *City* level in both years. Moreover, this trend is stronger in 2010 than in 2000. Similar results hold at the *Agglomeration* level and even at the *Region* level, the only exceptions in 2010 are the coefficients for *College* and in 2000 for *Highschool* and *College*. To summarize, the elasticity test provides strong support for hypothesis 1 that larger locations are relatively more skill abundant. This holds for both 2000 and 2010, but the results are stronger for 2010 than for 2000.

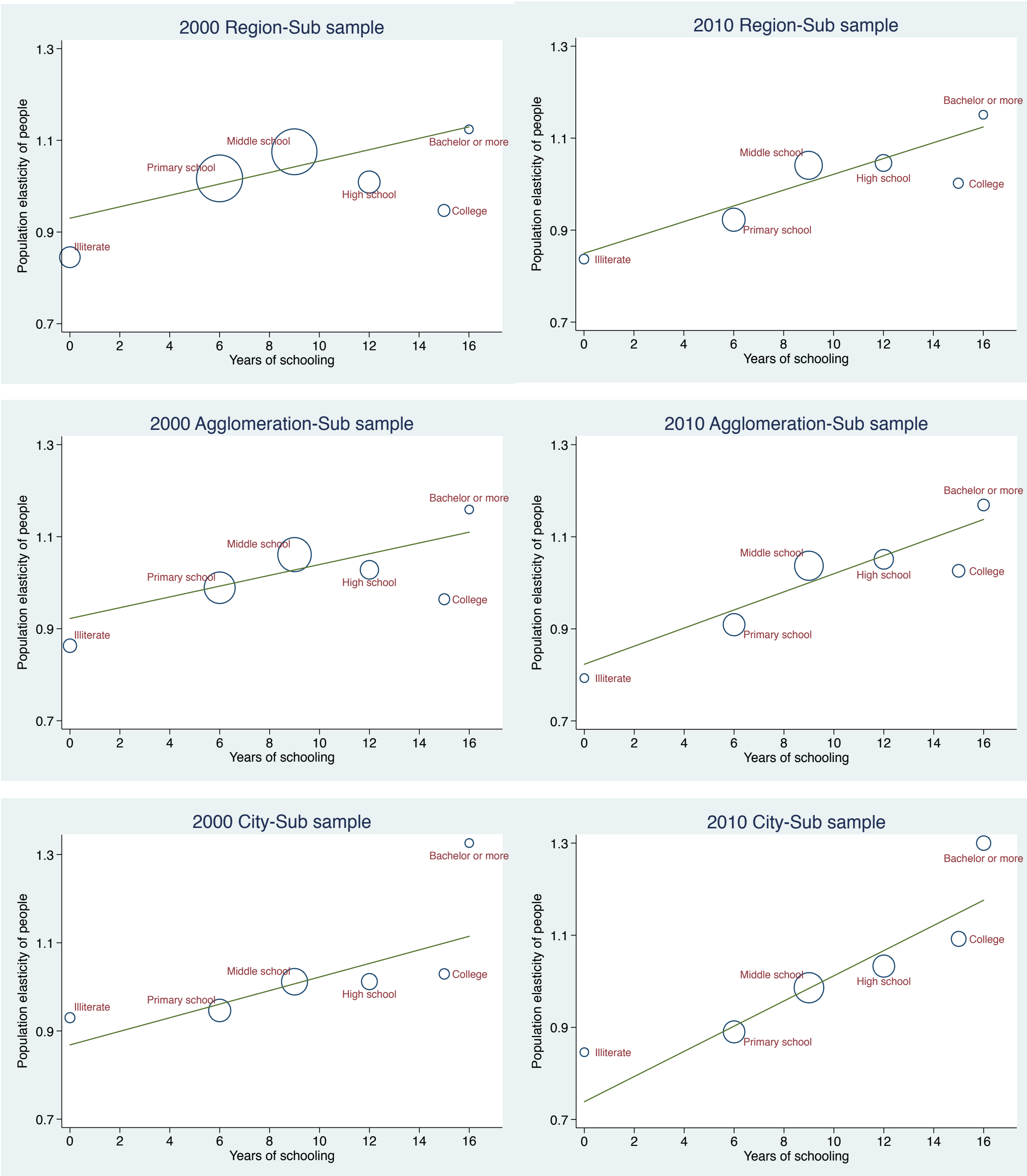
Table 4.5: Population elasticities of educational groups

Educational attainment	Region		Agglomeration		City	
	2000	2010	2000	2010	2000	2010
	(1)	(2)	(3)	(4)	(5)	(6)
Illiterate	0.845	0.837	0.863	0.793	0.930	0.846
	(0.049)	(0.057)	(0.031)	(0.033)	(0.035)	(0.039)
Primary school	1.017	0.923	0.989	0.909	0.946	0.890
	(0.042)	(0.039)	(0.033)	(0.026)	(0.028)	(0.028)
Middle school	1.075	1.041	1.061	1.037	1.012	0.986
	(0.072)	(0.071)	(0.043)	(0.042)	(0.040)	(0.041)
High school	1.009	1.046	1.028	1.051	1.012	1.033
	(0.090)	(0.083)	(0.046)	(0.044)	(0.055)	(0.051)
College	0.947	1.002	0.964	1.026	1.029	1.092
	(0.088)	(0.084)	(0.049)	(0.048)	(0.062)	(0.056)
Bachelor or more	1.124	1.151	1.159	1.169	1.326	1.300
	(0.117)	(0.105)	(0.066)	(0.062)	(0.074)	(0.067)
Observations	1,776	1,776	1,692	1,722	1,506	1,626
R-squared	0.909	0.893	0.911	0.911	0.889	0.899
Education FE	Yes	Yes	Yes	Yes	Yes	Yes

Note: standard errors in parentheses, clustered by prefectural cities.

To illustrate our findings we graph the population elasticities of the six educational groups listed in Table 4.5 relative to the corresponding educational levels in both years in Figure 4.2. We do this for all three location levels in a bubble diagram, where the size of the bubble is proportional to the population share of that education level. It is clear that educational levels *Middle school* and *Primary school* account for the largest proportion of the total population at the *Region* level in 2000, while the composition of educational attainment is more balanced at the *City* level in the same year. The *Agglomeration* level is intermediate of these two extremes. This implies that the areas with more urban features are more skill abundant. The educational composition was more balanced in 2010 for all thee location levels (with *Middle school* as the largest group), implying that the gap between rural and urban areas was getting smaller in this time period. The diagrams also display a regression line (weighted by population shares) for the estimated elasticities relative to the years of schooling. Based on this the fit relative to the model seemed to improve over time for all three location levels as the bubbles get closer to the fitted line and the slopes of the fitted lines are steeper in 2010, especially for *Agglomeration* and *City*.

Figure 4.2: Population elasticities of skills and years of schooling



Note: The size of the bubble measures the size of each educational level; The fitted lines are weighted by population shares; The vertical axis does not start at zero; The ‘sub-sample’ excludes the remote provinces Xinjiang, Tibet, Qinghai, and Inner Mongolia, see the appendix for results on the full sample.

If we look at the geographic dimension of the trend for the three types of locations we see, as expected, that the model performs better at the *Agglomeration* level than at the *Region* level and better at the *City* level than at the *Agglomeration* level. This holds for both time periods, but particularly for 2010. In this respect it is worth noting that the elasticities for

Bachelor or more are positive outliers at the *City* level in both years, implying that people with the highest education levels choose to live in larger cities.

Table 4.6: Success rate of hypothesis 1 elasticity test: large locations are more skill intensive

		Region		Agglomeration		City	
Year	Pairs	Rejection	Success (%)	Rejection	Success (%)	Rejection	Success (%)
2000	15	3	80.0	2	86.7	0	100.0
2010		0	100.0	0	100.0	0	100.0

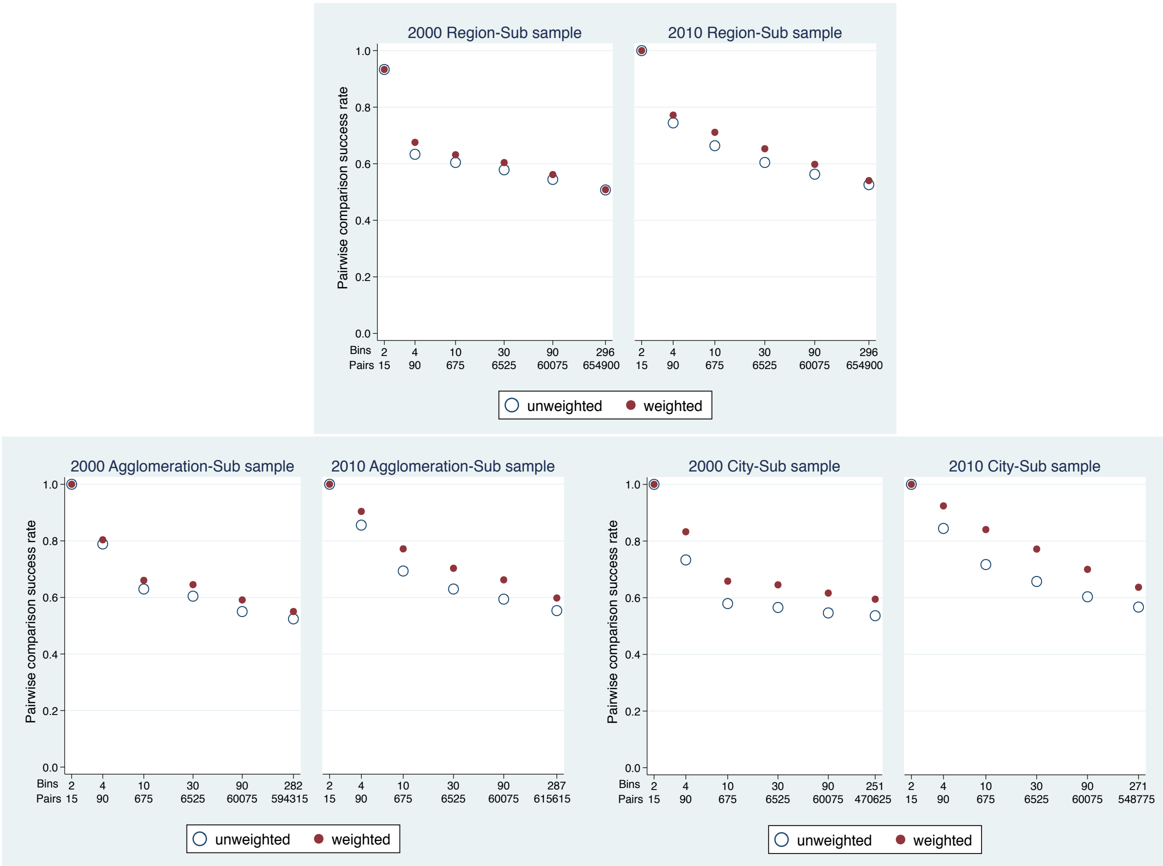
Success rate = 100%*(Pairs-Rejection)/Pairs; see main text for details.

Table 4.6 provides a formal test of the hypothesis that the estimated elasticities rise with higher education levels. More formally, the hypothesis that $\beta_{s3} \geq \beta_{s'3} \leftrightarrow s \geq s'$ involves 15 (=6*5/2) comparisons of the population elasticities in six educational groups. *Rejection* reports the number of comparisons that reject this hypothesis at the five-percent significance level. Taking the test for the *Region* level in 2000 as an example, this hypothesis is rejected in 3 out of 15 comparisons, resulting in a success rate of 80 percent. It is clear that the success rates increase over time. By 2010, *all* the successes were 100 percent in all location types. The success rate improved over time for *Region* and *Agglomeration*, while the success rate is 100 percent at the City level in both years.

B. Pairwise comparison test

Next, we focus on the pairwise comparison test regarding the relationship between location size and skill abundance. As explained in the previous section, by examining 'bins' of ordered groups of cities, the pairwise comparison test examines whether the relatively more skilled population is to be found in relatively large locations. Since we analyze 2, 4, 10, 30, and 90 bins as well as 296 individual locations (*Region* level) for 6 different skill categories, we make 722,280 bilateral comparisons for each location type for each year. The results are summarized in Figure 4.3 both regarding the unweighted and weighted success rate of the pairwise comparison tests (consisting in total of about 4.3 million bilateral comparisons, see Appendix 4.12 for details).

Figure 4.3: Pairwise comparison of six educational attainment levels, subsample



As expected, the success rates of these tests were higher in 2010 than that in 2000 for all three types of locations. As with the elasticities test, the geographic differences are clear. The success rate is highest for *City*, followed by *Agglomeration*, followed by *Region*. Restricting attention to the urban areas improves results significantly. It is also clear that the success rate of the pairwise comparison tests improves if we lump cities together in bigger groups (and thus a lower number of bins). The smallest groups of individual cities have a weighted succes rate ranging from 50 percent (for *Regions* in 2000) to above 60 percent (for *Cities* in 2010). In contrast, the succes rate when we have only two bins (containing half of the sample per bin) is 100 percent (with the exception of *Regions* in 2000). The weighted success rates are higher than the unweighted ones, indicating that the comparison test is more likely to hold if the difference in the size of the populations of the compared locations is big. Note that the gap between the weighted and unweighted results is largest for *City* and smallest for *Region*.

Our findings above suggest that the theoretical model works quite well regarding the relationship between location size and skill abundance. As expected, the tests perform better as we go from the *Region* level to the *Agglomeration* level, and again as we go from the *Agglomeration* level to the *City* level. This suggests that the model is more appropriate if the locational scale is more precisely and more coherently defined. In addition, the tests perform better in 2010 than in 2000. We take this as an indication of China's move over time to a more market-oriented economy allowing for greater labor mobility.[71]

4.5.2 Larger cities specialized in skill-intensive sectors and occupations

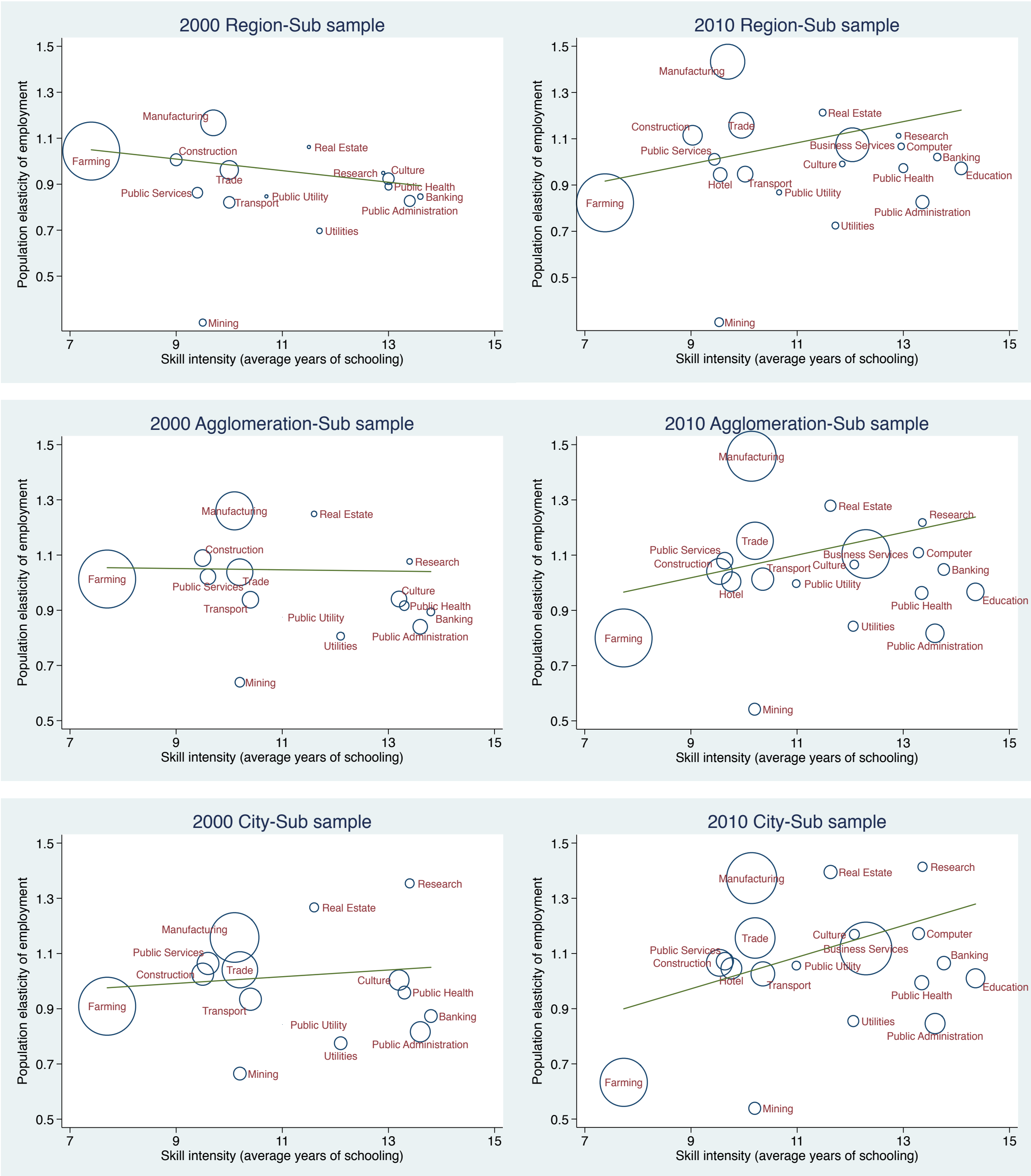
In this subsection, we tests whether larger cities are more specialized in highly skill-intensive sectors and occupations in 2000 and 2010. First, we estimate the sectoral and occupational population elasticities using the elasticity test. Second, we use the pairwise comparison test to identify the spatial patterns of sectoral and occupational employment.

A. Elasticity test-sectors

We start with the elasticity test for sectoral employment. As mentioned in subsection 4.4.3, the population elasticities at the *Region* level are based on the *Total* skill intensities, while the population elasticities at the *Agglomeration* and *City* levels are based on the *Urban* skill intensities. Figure 4.4 plots the 15 and 19 sectoral population elasticities against their corresponding skill intensities in 2000 and 2010. In general, the sectoral composition, measured by the size of the bubbles, shows a diversified pattern among the three location levels. In particular, *Farming* dominates at the *Region* level, particularly in 2000, while *Manufacturing* becomes more important for the urban areas, also over time. Also note that the sector *Business Services*, which was absent in the sectoral categories in 2000, accounts for the largest proportion of the total employment at the *City* level in 2010.

[71] Similar, but somewhat weaker, results hold when we analyze the full sample (including the remote provinces Xinjiang, Tibet, Qinghai, and Inner Mongolia), see the Appendix 4.3 and 4.4.

Figure 4.4: Sectors' population elasticities and skill intensities, subsample



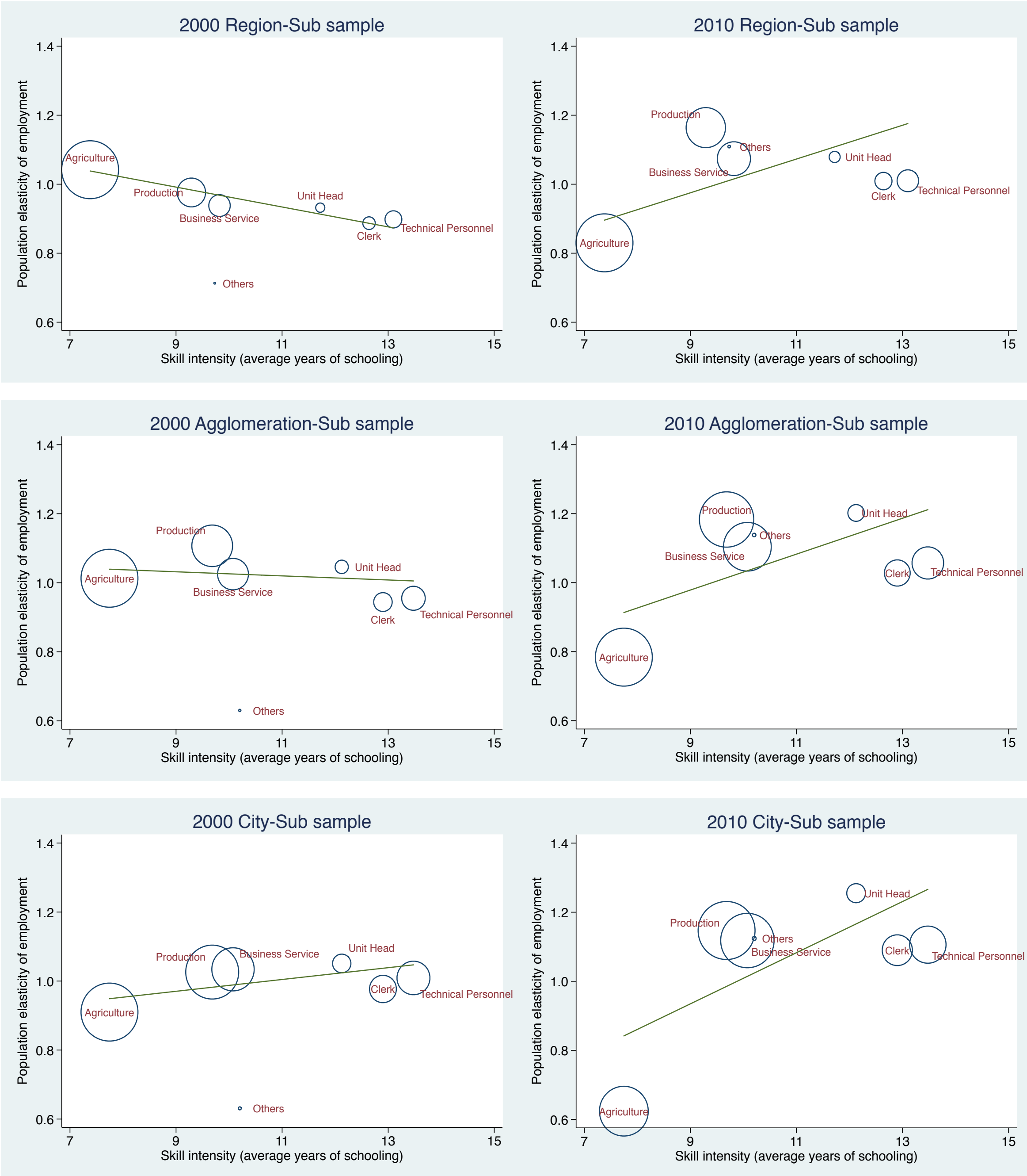
Note: The size of the bubble measures the size of sectors. The fitted lines are weighted by population shares. The vertical axis does not start at zero.

The elasticity test performs badly in 2000. The fitted line is basically horizontal for *Agglomeration* and *City*, while it even has a negative slope for *Region* (see Figure 4.4, left). The dominant high estimated elasticity for *Farming* at the *Region* level is the main culprit, indicating that large rural areas with a big population also have a lot of farming. For the urban areas *Real Estate* and *Research* have high elasticities, while *Mining*, *Public Utilities*, *Utilities*, *Public Health*, *Public Administration*, and *Banking* have low elasticities relative to their average schooling levels. The test performance improves by 2010. All fitted lines at least have a positive slope, indicating that by then large locations produce relatively more in skill-intensive sectors (see Figure 4.4, right). The estimated elasticity for low-skilled *Farming* falls considerably for all locations, particularly for *Cities*. Perhaps this reflects the decline of hidden unemployment in rural areas. A summary of the formal tests of the comparison of the various estimated elasticities is reported in Table 4.7 below. The test of the hypothesis that $\beta_{\sigma 3} \geq \beta_{\sigma' 3} \leftrightarrow \sigma \geq \sigma'$ at the sector level requires 105 (=15*14/2) and 171 (19*18/2) comparisons in 2000 and 2010, respectively. Based on the fairly low succes rate of the elasticity test, which is always below 64 percent and only improves somewhat over time, we cannot really conclude that high-skilled sectors are concentrated in larger locations. Perhaps the best conclusion is that the sectoral composition of skills is too diversified to become clearly visible at the sorting stage in China at this point in time.

B. Elasticity test-occupations

In this subsection, we use the elasticity test to examine whether large locations specialize in relatively high-skilled occupations. Our data identifies only seven occupational categories. The estimated population elasticities are plotted in Figure 4.5 relative to the average number of years of schooling in 2000 and 2010 for the three location levels, with the size of the bubbles proportional to the number of people working in that occupation. As with the sector analysis, *Agriculture* takes up a dominant position in size, particularly at the *Region* level and particularly in 2000. The share of employment in *Production* and *Business Services* increased substantially over time (they became the two largest occupation categories in 2010 at the urban levels). As with the sector analysis, the slope of the fitted lines for the elasticities of occupations is basically horizontal in 2000 for *Agglomeration* and even negative for *Regions*. As with the sector analysis, the predictions perform better in 2010 with fitted lines with a positive slope for all three location types. As with the sector analysis, the estimated elasticity of *Agriculture* as an occupation declined substantially in 2010. In contrast to the sector analysis, however, a visual inspection suggests that the overall performance of the elasticity test seems to be quite acceptable by 2010, see also the formal analysis discussed below.

Figure 4.5: Occupations' population elasticities and skill intensities, subsample



Note: The size of the bubble measures the size of occupations. The fitted lines are weighted by population shares. The vertical axis does not start at zero.

Table 4.7 (occupational part) shows the formal test for the population elasticities for occupations. The test of the hypothesis that $\beta_{\sigma 3} \geq \beta_{\sigma' 3} \leftrightarrow \sigma \geq \sigma'$ for occupations requires 21 (=7*6/2) comparisons of the estimated elasticities in both years. The succes rate is highest in 2000 at the *Region* level (because many differences are not statistically significant), but it falls over time to become the lowest in 2010. For the urban levels *Agglomeration* and *City* the success rate improves over time. It remains relatively low in

2010 at the *Agglomeration* level (67 percent), but becomes high at the *City* level (more than 90 percent). We can therefore conclude that higher skilled occupations are indeed concentrated in larger cities by 2010. As expected, most of the success rates for occupations are higher than for sectors, suggesting that the model is more appropriate at the occupation level than at the sector level.

Table 4.7: Success rate of hypothesis 2

			Region		**Agglomeration**		**City**	
	Year	Pairs	Rejection	Success (%)	Rejection	Success (%)	Rejection	Success (%)
Sector	2000	105	38	63.8	51	51.4	43	59.0
	2010	171	63	63.2	70	59.1	62	63.7
Occupation	2000	21	4	81.0	11	47.6	5	76.2
	2010	21	8	61.9	7	66.7	2	90.5

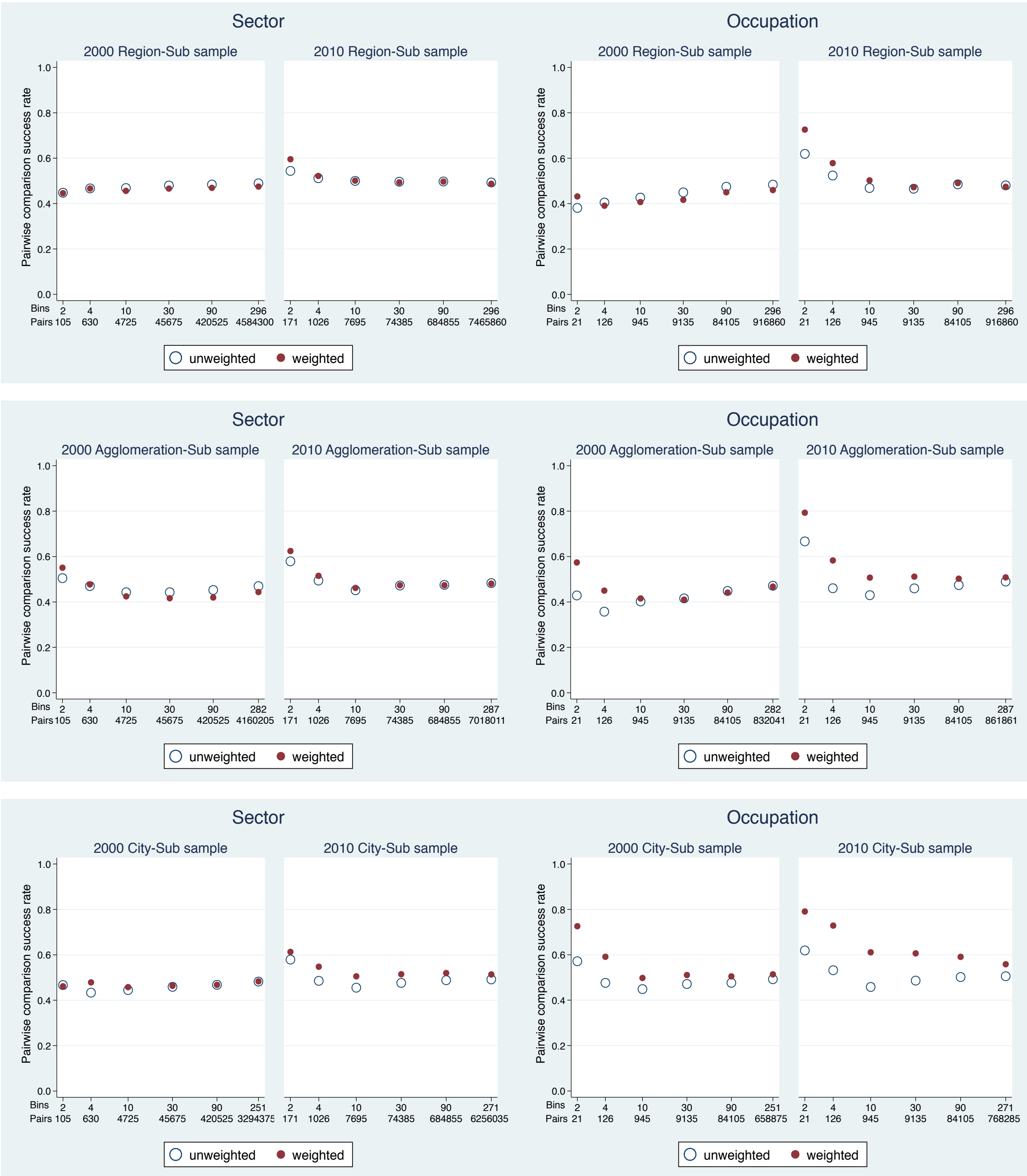
C. Pairwise comparison test

We next turn to the results of the pairwise comparison test for sectors and occupations (see Figure 4.6, the Appendix 4.13 and 4.14). Since we have 7 different occupation categories, each figure for a given location level and year for occupations is based on about 1 million bilateral comparisons. Since we have 15 sectors in 2000 and 19 sectors in 2010, each figure for a given location level and year for sectors is based on about 5 million bilateral comparisons in 2000 and 8.2 million bilateral comparisons in 2010. In total, Figure 4.6 is thus based on about 46 million bilateral comparisons, both weighted and unweighted.

At the *sector* level, the success rate of the bilateral comparisons improves from 2000 to 2010 for all three types of locations. This holds in particular for the weighted values of the comparison of larger groups of cities (the lower bin numbers). Nonetheless, even for those groups the success rate remains small in 2010, only around 60 percent. The majority of the comparisons (either weighted or unweighted) is close to the 50 percent benchmark.

At the *occupation* level, the success rate of the bilateral comparisons also improves from 2000 to 2010. This time, however, the improvement is more substantial. This holds in particular for the weighted success rates, for the larger groups of cities comparisons (success rates of 80 percent for the lower bin numbers), and for the *City* level. This confirms our earlier analysis that performance improves over time, is better for more consistently defined locations, and that the model is more appropriate for occupations than for sectors.

Figure 4.6: Pairwise comparison of 19 (15) sectors and 7 occupations, subsample



To summarize, we find some evidence that larger *cities* in China have become relatively more specialized in skill-intensive sectors and occupations over the past decade. This change may be indicative of the more market-oriented economic development accompanied by more labour mobility and rapid urbanization from 2000 to 2010. The degree of urbanization in China in 2000 was only about 37 percent, indicating that about two-thirds of the population still lived in rural areas and relied primarily on farming.[72] Since *Region* and *Agglomeration* in China include not only urban areas but also rural areas, larger cities are related to more agricultural employment than smaller cities. As agricultural activities belong to low skill-intensive economies, the estimated results do not confirm our predictions in *Region* and *Agglomeration* in 2000, for both sectors and occupations. In contrast, the degree of urbanization has increased to about 50 percent in 2010. The move to a more market-oriented economy based on manufacturing and business services attracted a large number of workers from rural into the urban areas. This transition is illustrated by the improvement in predictive power over time of the model as tested in this section.

4.6 Conclusions

A large urbanization wave is surging in China. Millions of jobs in offices, factories, shopping malls, and construction sites are created in the urban areas, attracting more and more people from the countryside. This rural-urban migrant flow largely shaped the reallocation of labor, industry and wealth, forming the foundation of modern China. We empirically test the theoretical framework of Davis and Dingel (2013) which predicts that there is relative sorting of high-skilled workers in larger cities. Associated with this process there is also a relative sorting of high-skilled occupations and skill-intensive sectors in larger cities.

Our analysis is based on two tests, namely the population elasticity test and the pairwise comparison test. We do this for three types of Chinese locations (regions, agglomerations, and cities) and for three types of observables (skills, occupations, and sectors) in 2000 and 2010. The elasticity test holds if the estimated population elasticity is higher for higher skills (or for more skill-intensive occupations or sectors). The pairwise comparison test

[72] The data of urbanization is calculated from the Chinese census of population. Urbanization is defined as the share of the population living in all districts and the center of towns below county levels in total population.

holds if the largest (group of) location(s) is relatively skill abundant (or skill-intensive occupation or sector abundant) if we compare two (groups of) cities. In all cases, the results found by the elasticity test are in line with the results found by the pairwise comparison test. Our main findings can be summarized as follows.

- The predictive power of the sorting model is highest for cities, followed by agglomerations and regions, respectively.
- The predictive power of the sorting model is highest for skills, followed by occupations and sectors, respectively.
- The predictive power of the sorting model improves over time.

We view our results as an indication that China's economy is transforming into a more market-oriented economy which allows for more labor mobility over time. We also think that our results indicate that care should be given regarding the type of location (the level of aggregation). In particular, the sorting model works best when comparing rather precisely defined and coherent locations (cities) rather than more diffuse and scattered locations (regions), as individuals base their location decisions on spatial interactions at the appropriate level only. Finally, we think that our results indicate that the sorting model works best when comparing skill levels as directly as possible, which explains the ranking for (education) skills, occupations, and sectors.

4.7 References

Abdel-Rahman, H. M. and Anas, A. (2004). Theories of systems of cities. *Handbook of Regional and Urban Economics*, *4*, 2293-2339.

Arzaghi, M. and Henderson, J. V. (2008). Networking off madison avenue. *The Review of Economic Studies*, *75*(4), 1011-1038.

Au, C. C. and Henderson, J. V. (2006). How migration restrictions limit agglomeration and productivity in China. *Journal of Development Economics*,*80*(2), 350-388.

Bacolod, M., Blum, B. S. and Strange, W. C. (2009). Skills in the City. *Journal of Urban Economics*, *65*(2), 136-153.

Behrens, K., Duranton, G., and Robert-Nicoud, F. (2013). Productive cities: Sorting, selection, and agglomeration (No. 13111). Département des Sciences Économiques, Université de Genève.

Chen, Y., Jin, G. Z. and Yue, Y. (2010). Peer Migration in China. *NBER Working Paper Series*, 15671.

Costinot, A. (2009). An elementary theory of comparative advantage. *Econometrica*, *77*(4), 1165-1192.

Davis, D. R., and Dingel, J. I. (2013). The comparative advantage of cities. Columbia University, mimeo.

Duranton, G. and Overman, H. G. (2005). Testing for localization using micro-geographic data. *The Review of Economic Studies*, *72*(4), 1077-1106.

Du, Y., Park, A. and Wang, S. (2005). Migration and rural poverty in China. *Journal of comparative economics*, *33*(4), 688-709.

Ellison, G. (1997). Geographic Concentration in US Manufacturing Industries: A Dartboard. *The Journal of Political Economy*, *105*(5), 889-927.

Ellison, G., Glaeser, E. L. and Kerr, W. R. (2010). What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns. *The American Economic Review*, *100*(3), 1195-1213.

Fu, Y. and Gabriel, S. A. (2012). Labor migration, human capital agglomeration and regional development in China. *Regional Science and Urban Economics*,*42*(3), 473-484.

Ge, Y. (2009). Globalization and industry agglomeration in China. *World development*, *37*(3), 550-559.

Glaeser, E. L. (1999). Learning in cities. *Journal of urban Economics*, *46*(2), 254-277.

Glaeser, E. L. and Resseger, M. G. (2010). The complementarity between cities and skills. *Journal of Regional Science*, *50*(1), 221-244.

Helsley, R. W. and Strange, W. C. (2012). Coagglomeration and the Scale and Composition of Clusters. UC Berkeley, mimeo

Henderson, J. V. (1974). The Sizes and Types of Cities. *The American Economic Review*, 64(4, 640-56.)

Henderson, J. V. (1983). Industrial bases and city sizes. *American Economic Review*, *73*(2), 164-68.

Henderson, V. (1997). Medium size cities. *Regional science and urban economics*, *27*(6), 583-612.

Holmes, T. J. (1998). The effect of state policies on the location of manufacturing: Evidence from state borders. *Journal of Political Economy*, *106*(4), 667-705.

Holmes, T. J. and Stevens, J. J. (2004). Spatial distribution of economic activities in North America. *Handbook of regional and urban economics*, *4*, 2797-2843.

Krugman, P. (1991). Increasing Returns and Economic Geography. *The Journal of Political Economy*, *99*(3), 483-499.

Liu, Z. (2005). Institution and inequality: the *hukou* system in China. *Journal of Comparative Economics*, *33*(1), 133-157.

Marshall, A. (1920). *Principles of Economics*. London: Macmillan and Co., Ltd.

Milgrom, P. (1981). Good news and bad news: Representation theorems and applications. *Bell journal of Economics*, *12*(2), 380-391.

Mori, T. and Turrini, A. (2005). Skills, agglomeration and segmentation. *European Economic Review*, *49*(1), 201-225.

Wen, M. (2004). Relocation and agglomeration of Chinese industry. *Journal of Development Economics*, *73*(1), 329-347.

Whalley, J. and Zhang, S. (2007). A numerical simulation analysis of (Hukou) labour mobility restrictions in China. *Journal of Development Economics*, *83*(2), 392-410.

Young, A. (2003). Gold into base metals: productivity growth in the People's Republic of China during the Reform Period. *Journal of political economy*,*111*(1), 1220-1261.

Zhang, K. H. and Song, S. (2003). Rural–urban migration and urbanization in China: Evidence from time-series and cross-section analyses. *China Economic Review*, *14*(4), 386-400.

Zhao, Y. (2002). Causes and consequences of return migration: recent evidence from China. *Journal of Comparative Economics*, *30*(2), 376-394.

Zhu, N. and Luo, X. (2010). The impact of migration on rural poverty and inequality: a case study in China. *Agricultural Economics*, *41*(2), 191-204

4.8 Appendix

Appendix 4.1: Full name and short name of sector and occupation

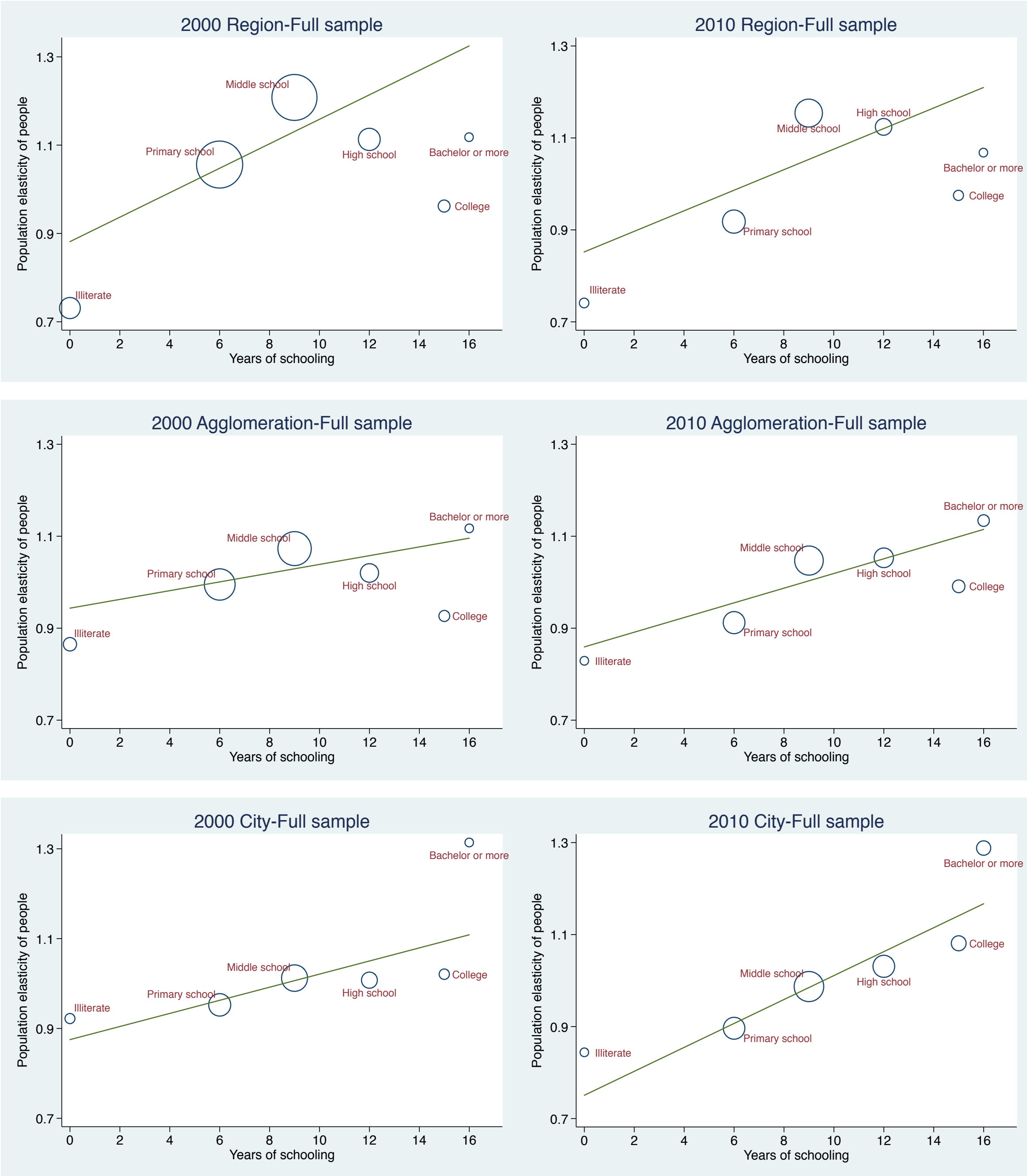
A. Sector		
Short name	**Full name-2000**	**Full name-2010**
Farming	Farming, Forestry, Animal Husbandry and Fishery	Farming, Forestry, Animal Husbandry and Fishery
Construction	Construction	Construction
Public Services	Social Service	Personal and other Services
Mining	Mining and Quarrying	Mining and Quarrying
Hotel	-	Hotel and Catering Services
Manufacturing	Manufacturing	Manufacturing
Trade	Wholesale and Retail Trade & Catering Services	Wholesale and Retail Trade
Transport	Transport, Storage, Post & Telecommunications	Transport, Storage, Post & Telecommunications
Public Utility	Geological Prospecting & Water Conservancy	Water Conservancy, Environment and public Utility Management
Real Estate	Real Estate Trade	Real Estate
Utilities	Production and Supply of Electricity Gas and Water	Production and Supply of Electricity Gas and Water
Culture	Education, Culture and Art, Radio Film and Television	Culture Sports and Entertainment
Business Services	-	Leasing and Business Services
Research	Scientific Research and Poly-technical Services	Scientific Research, Technical Services & Geological Prospecting
Computer	-	Data Transmission Computer Service & Software
Public Health	Health Care, Sporting and Social Welfare	Public Health Social Securities & Social Welfare
Public Administration	Government Agencies, Party Agencies and Social Organizations	Public Administration and Social Organizations
Banking	Finance and Insurance	Banking
Education	-	Education
B. Occupation		
Short names	**Full name-2000**	**Full name-2010**
Agriculture	Agriculture and Water Conservancy Laborers	Agriculture and Water Conservancy Laborers
Production	Production, Transport Equipment Operators and Related Workers	Production, Transport Equipment Operators and Related Workers
Others	Others	Others
Business Service	Business Service Personnel	Business Service Personnel
Unit Head	Unit Head	Unit Head
Clerk	Clerk and Related Workers	Clerk and Related Workers
Technical Personnel	Professional and Technical Personnel	Professional and Technical Personnel

Appendix 4.2: Population elasticities of educational groups, full sample

Educational attainment	Region		Agglomeration		City	
	2000	2010	2000	2010	2000	2010
	(1)	(2)	(3)	(4)	(5)	(6)
Illiterate	0.731	0.741	0.865	0.829	0.922	0.844
	(0.046)	(0.054)	(0.028)	(0.034)	(0.034)	(0.038)
Primary school	1.056	0.918	0.995	0.912	0.953	0.896
	(0.050)	(0.045)	(0.029)	(0.029)	(0.029)	(0.028)
Middle school	1.208	1.154	1.073	1.047	1.012	0.987
	(0.082)	(0.079)	(0.039)	(0.042)	(0.040)	(0.041)
High school	1.113	1.124	1.020	1.053	1.008	1.031
	(0.092)	(0.084)	(0.043)	(0.043)	(0.054)	(0.050)
College	0.962	0.975	0.927	0.991	1.021	1.081
	(0.085)	(0.078)	(0.047)	(0.049)	(0.060)	(0.055)
Bachelor or more	1.118	1.068	1.117	1.134	1.314	1.288
	(0.094)	(0.084)	(0.060)	(0.057)	(0.072)	(0.065)
Observations	2,028	2,022	1,872	1,896	1,572	1,704
R-squared	0.907	0.893	0.913	0.912	0.889	0.899
Education FE	Yes	Yes	Yes	Yes	Yes	Yes

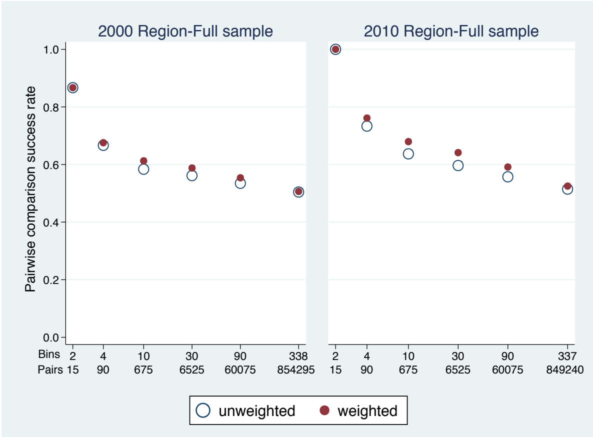
Note: standard errors in parentheses, clustered by prefectural cities.

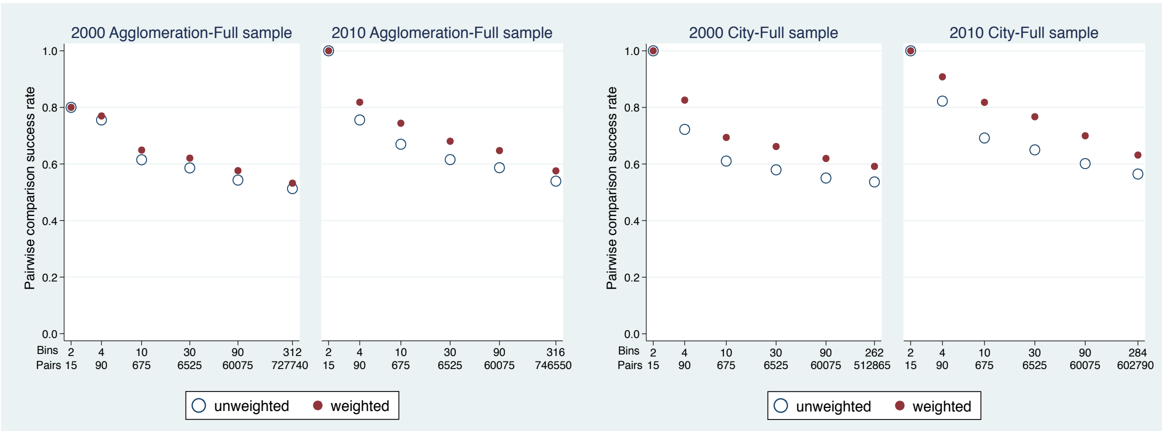
Appendix 4.3: Population elasticities of skills and years of schooling, full sample



Note: The size of the bubble measures the size of each educational level.
The vertical axis does not start at zero.

Appendix 4.4: Pairwise comparison of six educational attainment, full sample





Appendix 4.5: Sectoral employment population elasticities, 2000

	Region		Agglomeration		City	
Sector	Full	Sub	Full	Sub	Full	Sub
	(1)	(2)	(3)	(4)	(5)	(6)
Farming	1.041	1.044	1.037	1.013	0.946	0.909
	(0.044)	(0.059)	(0.051)	(0.058)	(0.073)	(0.071)
Construction	1.068	1.006	1.031	1.089	1.012	1.025
	(0.093)	(0.109)	(0.077)	(0.085)	(0.102)	(0.098)
Public Services	0.946	0.863	0.954	1.021	1.044	1.064
	(0.093)	(0.115)	(0.081)	(0.090)	(0.109)	(0.106)
Mining	0.633	0.298	0.703	0.639	0.654	0.665
	(0.128)	(0.134)	(0.118)	(0.130)	(0.145)	(0.145)
Manufacturing	1.276	1.167	1.245	1.260	1.156	1.158
	(0.098)	(0.124)	(0.080)	(0.092)	(0.107)	(0.107)
Trade	0.991	0.962	0.984	1.038	1.024	1.041
	(0.075)	(0.098)	(0.071)	(0.080)	(0.098)	(0.094)
Transport	0.903	0.821	0.895	0.938	0.914	0.934
	(0.083)	(0.102)	(0.069)	(0.075)	(0.095)	(0.090)
Public Utility	0.834	0.847	0.748	0.875	0.821	0.842
	(0.090)	(0.092)	(0.081)	(0.086)	(0.103)	(0.098)
Real Estate	1.122	1.062	1.271	1.249	1.256	1.267
	(0.143)	(0.182)	(0.109)	(0.121)	(0.128)	(0.128)
Utilities	0.782	0.697	0.795	0.806	0.768	0.775
	(0.082)	(0.098)	(0.076)	(0.084)	(0.098)	(0.096)
Research	0.884	0.949	1.016	1.077	1.312	1.354
	(0.108)	(0.151)	(0.104)	(0.113)	(0.130)	(0.122)
Culture	0.915	0.924	0.894	0.941	0.989	1.004
	(0.059)	(0.074)	(0.064)	(0.070)	(0.088)	(0.084)
Public Health	0.888	0.890	0.878	0.917	0.947	0.959
	(0.064)	(0.085)	(0.066)	(0.074)	(0.090)	(0.087)
Public	0.764	0.827	0.792	0.84	0.802	0.816
Administration	(0.054)	(0.075)	(0.066)	(0.072)	(0.093)	(0.089)
Banking	0.869	0.846	0.864	0.894	0.865	0.873
	(0.074)	(0.094)	(0.074)	(0.081)	(0.100)	(0.098)
Observations	5,062	4,439	4,679	4,229	3,929	3,764
R-squared	0.865	0.871	0.852	0.858	0.819	0.824
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes

Note: standard errors in parentheses, clustered by prefectural cities.
Full and Sub denote the full sample and sub sample.

Appendix 4.6: Sectoral employment population elasticities, 2010

	Region		**Agglomeration**		**City**	
Sector	Full	Sub	Full	Sub	Full	Sub
	(1)	(2)	(3)	(4)	(5)	(6)
Farming	0.868	0.823	0.827	0.800	0.662	0.633
	(0.058)	(0.084)	(0.064)	(0.072)	(0.076)	(0.076)
Construction	1.216	1.115	1.043	1.041	1.054	1.066
	(0.086)	(0.099)	(0.075)	(0.080)	(0.080)	(0.079)
Public Services	1.103	1.011	1.090	1.080	1.065	1.072
	(0.086)	(0.113)	(0.080)	(0.087)	(0.089)	(0.089)
Mining	0.634	0.307	0.617	0.542	0.541	0.539
	(0.131)	(0.121)	(0.111)	(0.117)	(0.114)	(0.114)
Hotel	1.009	0.945	1.015	1.004	1.035	1.046
	(0.082)	(0.112)	(0.082)	(0.090)	(0.092)	(0.092)
Manufacturing	1.55	1.433	1.462	1.458	1.368	1.373
	(0.109)	(0.145)	(0.088)	(0.098)	(0.105)	(0.108)
Trade	1.176	1.158	1.137	1.152	1.147	1.156
	(0.080)	(0.112)	(0.077)	(0.084)	(0.089)	(0.089)
Transport	1.041	0.947	0.994	1.013	1.016	1.026
	(0.091)	(0.119)	(0.076)	(0.084)	(0.088)	(0.088)
Public Utility	0.922	0.868	0.960	0.997	1.040	1.056
	(0.101)	(0.126)	(0.088)	(0.098)	(0.097)	(0.097)
Real Estate	1.268	1.213	1.266	1.279	1.363	1.395
	(0.144)	(0.171)	(0.102)	(0.107)	(0.104)	(0.103)
Utilities	0.851	0.725	0.845	0.843	0.852	0.855
	(0.095)	(0.116)	(0.077)	(0.085)	(0.085)	(0.085)
Culture	0.968	0.991	1.029	1.066	1.147	1.169
	(0.098)	(0.137)	(0.094)	(0.103)	(0.103)	(0.102)
Business	1.114	1.074	1.094	1.104	1.108	1.118
Services	(0.082)	(0.114)	(0.078)	(0.086)	(0.089)	(0.089)
Research	1.055	1.113	1.111	1.218	1.380	1.414
	(0.119)	(0.159)	(0.109)	(0.117)	(0.117)	(0.114)
Computer	1.067	1.067	1.079	1.109	1.158	1.172
	(0.094)	(0.131)	(0.091)	(0.102)	(0.102)	(0.103)
Public Health	0.974	0.973	0.947	0.963	0.993	0.994
	(0.074)	(0.103)	(0.072)	(0.082)	(0.083)	(0.084)
Public	0.748	0.826	0.789	0.817	0.837	0.846
Administration	(0.071)	(0.100)	(0.080)	(0.091)	(0.092)	(0.092)
Banking	1.039	1.021	1.028	1.048	1.057	1.065
	(0.092)	(0.127)	(0.084)	(0.094)	(0.097)	(0.098)
Education	0.970	0.972	0.946	0.967	1.007	1.010
	(0.068)	(0.095)	(0.070)	(0.079)	(0.080)	(0.081)
Observations	6,397	5,624	6,003	5,453	5,396	5,149
R-squared	0.871	0.879	0.875	0.880	0.857	0.862
Sector FE	Yes	Yes	Yes	Yes	Yes	Yes

Note: standard errors in parentheses, clustered by prefectural cities.
Full and Sub denote the full sample and sub sample.

Appendix 4.7: Occupational employment population elasticities, 2000

	Region		**Agglomeration**		**City**	
Occupation	Full	Sub	Full	Sub	Full	Sub
	(1)	(2)	(3)	(4)	(5)	(6)
Agriculture	1.041	1.042	1.039	1.013	0.947	0.910
	(0.043)	(0.058)	(0.051)	(0.058)	(0.073)	(0.071)
Production	1.079	0.976	1.073	1.107	1.014	1.026
	(0.093)	(0.114)	(0.078)	(0.088)	(0.106)	(0.104)
Others	0.775	0.713	0.736	0.630	0.643	0.631
	(0.076)	(0.108)	(0.090)	(0.098)	(0.118)	(0.123)
Business	0.976	0.938	0.976	1.025	1.016	1.034
Service	(0.076)	(0.099)	(0.072)	(0.080)	(0.098)	(0.094)
Unit Head	0.871	0.932	0.992	1.046	1.033	1.051
	(0.074)	(0.100)	(0.076)	(0.084)	(0.104)	(0.101)
Clerk	0.888	0.887	0.905	0.944	0.964	0.977
	(0.073)	(0.100)	(0.077)	(0.087)	(0.105)	(0.103)
Technical	0.880	0.898	0.908	0.955	0.995	1.009
Personnel	(0.063)	(0.084)	(0.069)	(0.077)	(0.094)	(0.091)
Observations	2,355	2,062	2,168	1,960	1,815	1,738
R-squared	0.923	0.923	0.915	0.915	0.899	0.899
Occup FE	Yes	Yes	Yes	Yes	Yes	Yes

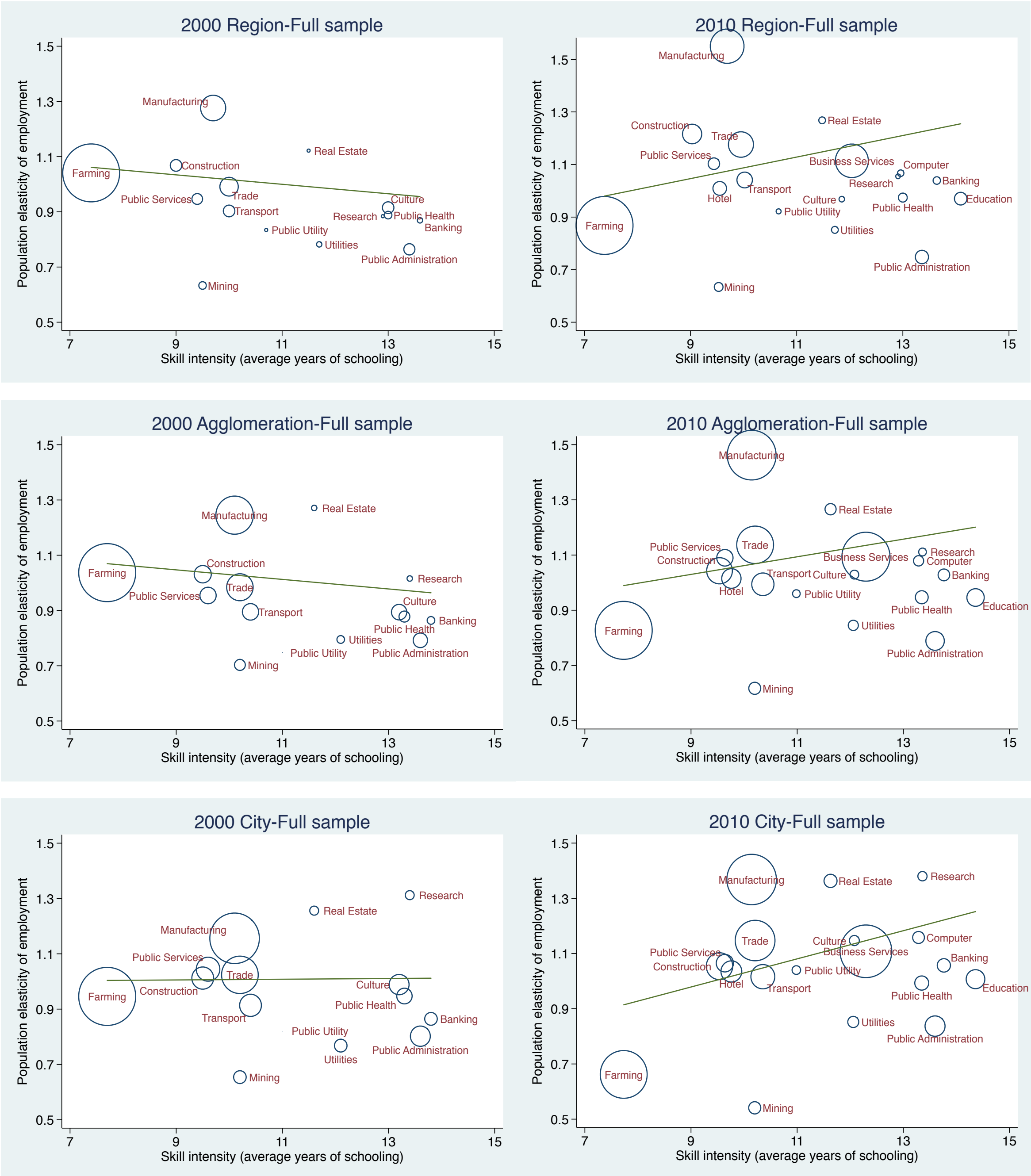
Note: standard errors in parentheses, clustered by prefectural cities
Full and Sub denote the full sample and sub sample.

Appendix 4.8: Occupational employment population elasticities, 2010

	Region		**Agglomeration**		**City**	
Occupation	Full	Sub	Full	Sub	Full	Sub
	(1)	(2)	(3)	(4)	(5)	(6)
Agriculture	0.870	0.830	0.817	0.784	0.654	0.624
	(0.058)	(0.084)	(0.061)	(0.068)	(0.073)	(0.073)
Production	1.277	1.164	1.191	1.183	1.136	1.147
	(0.098)	(0.124)	(0.081)	(0.088)	(0.093)	(0.094)
Others	1.142	1.109	1.105	1.138	1.149	1.124
	(0.102)	(0.145)	(0.111)	(0.117)	(0.135)	(0.136)
Business	1.113	1.074	1.094	1.104	1.109	1.118
Service	(0.082)	(0.114)	(0.076)	(0.082)	(0.087)	(0.086)
Unit Head	1.033	1.079	1.165	1.202	1.239	1.255
	(0.094)	(0.134)	(0.087)	(0.097)	(0.098)	(0.098)
Clerk	0.971	1.009	1.003	1.028	1.075	1.090
	(0.088)	(0.125)	(0.087)	(0.097)	(0.099)	(0.099)
Technical	0.983	1.010	1.018	1.057	1.097	1.106
Personnel	(0.080)	(0.113)	(0.078)	(0.086)	(0.089)	(0.089)
Observations	2,352	2,069	2,200	2,000	1,978	1,887
R-squared	0.922	0.922	0.912	0.916	0.899	0.901
Occup FE	Yes	Yes	Yes	Yes	Yes	Yes

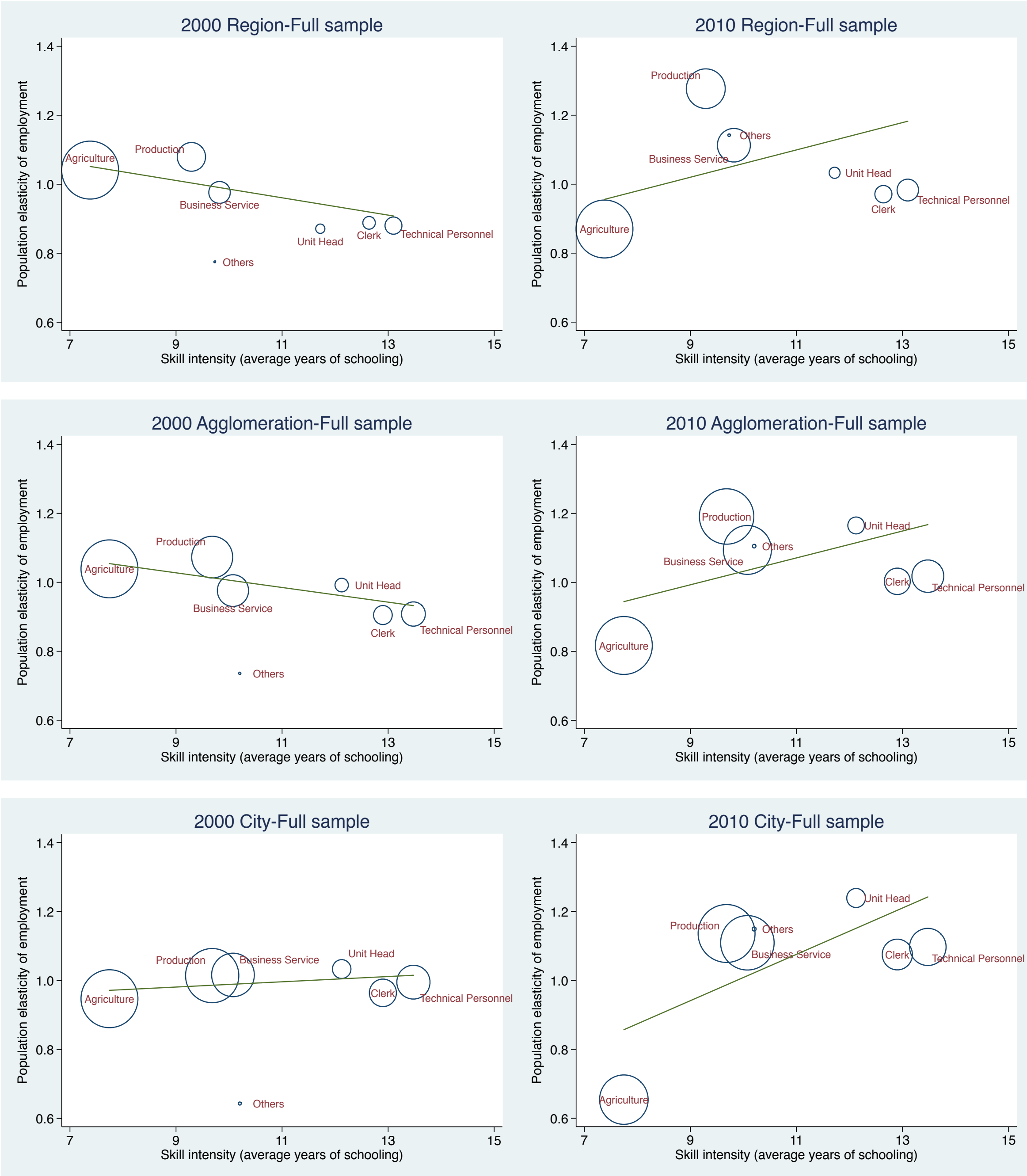
Note: standard errors in parentheses, clustered by prefectural cities
Full and Sub denote the full sample and sub sample.

Appendix 4.9: Sectors' population elasticities and skill intensities, full sample



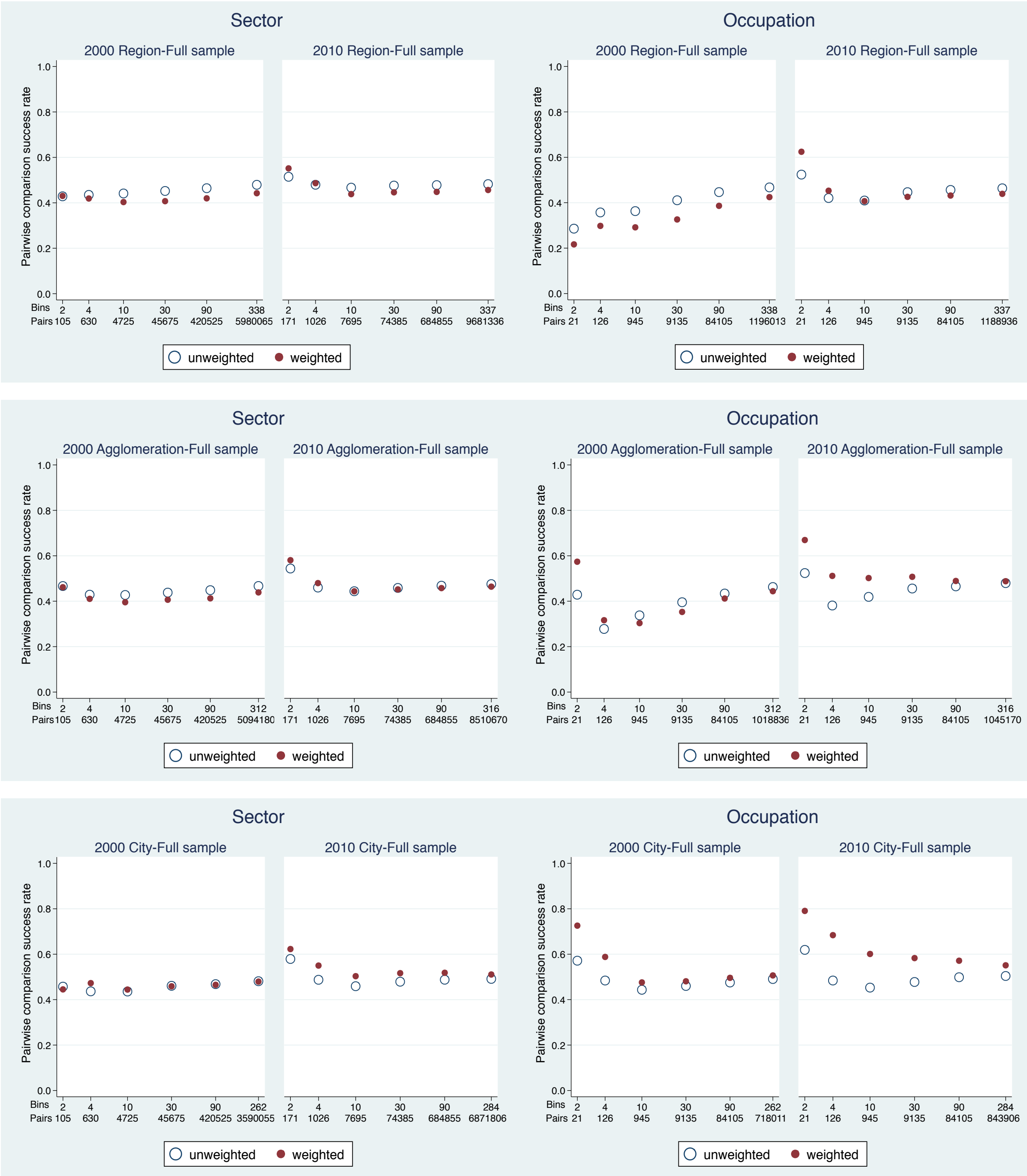
Note: The size of the bubble measures the size of sectors.
The vertical axis does not start at zero.

Appendix 4.10: Occupations' population elasticities and skill intensities, full sample



Note: The size of the bubble measures the size of occupations.
The vertical axis does not start at zero.

Appendix 4.11: Pairwise comparison of 19 sectoral and 7 occupational categories, full sample



Appendix 4.12: The pairwise comparison for skills

2000-Region					**2010-Region**				
Bin	**Pairs**	**Sample**	**Unweighted**	**Weighted**	**Bin**	**Pairs**	**Sample**	**Unweighted**	**Weighted**
2	15	All	0.867	0.867	2	15	All	1.000	1.000
2	15	Sub	0.933	0.933	2	15	Sub	1.000	1.000
4	90	All	0.667	0.676	4	90	All	0.733	0.761
4	90	Sub	0.633	0.676	4	90	Sub	0.744	0.772
10	675	All	0.584	0.613	10	675	All	0.637	0.679
10	675	Sub	0.604	0.632	10	675	Sub	0.664	0.711
30	6525	All	0.562	0.588	30	6525	All	0.596	0.641
30	6525	Sub	0.579	0.604	30	6525	Sub	0.604	0.653
90	60075	All	0.535	0.554	90	60075	All	0.557	0.591
90	60075	Sub	0.545	0.562	90	60075	Sub	0.563	0.598
338	854295	All	0.504	0.506	337	849240	All	0.515	0.525
296	654900	Sub	0.507	0.508	296	654900	Sub	0.526	0.540
2000-Agglomeration					**2010-Agglomeration**				
Bin	**Pairs**	**Sample**	**Unweighted**	**Weighted**	**Bin**	**Pairs**	**Sample**	**Unweighted**	**Weighted**
2	15	All	0.800	0.800	2	15	All	1.000	1.000
2	15	Sub	1.000	1.000	2	15	Sub	1.000	1.000
4	90	All	0.756	0.770	4	90	All	0.756	0.818
4	90	Sub	0.789	0.804	4	90	Sub	0.856	0.904
10	675	All	0.615	0.649	10	675	All	0.670	0.744
10	675	Sub	0.630	0.660	10	675	Sub	0.693	0.772
30	6525	All	0.586	0.621	30	6525	All	0.615	0.680
30	6525	Sub	0.604	0.645	30	6525	Sub	0.630	0.703
90	60075	All	0.543	0.576	90	60075	All	0.587	0.647
90	60075	Sub	0.550	0.591	90	60075	Sub	0.594	0.662
312	727740	All	0.513	0.532	316	746550	All	0.539	0.576
282	594315	Sub	0.524	0.550	287	615615	Sub	0.553	0.598
2000-City					**2010-City**				
Bin	**Pairs**	**Sample**	**Unweighted**	**Weighted**	**Bin**	**Pairs**	**Sample**	**Unweighted**	**Weighted**
2	15	All	1.000	1.000	2	15	All	1.000	1.000
2	15	Sub	1.000	1.000	2	15	Sub	1.000	1.000
4	90	All	0.722	0.826	4	90	All	0.822	0.908
4	90	Sub	0.733	0.833	4	90	Sub	0.844	0.924
10	675	All	0.610	0.694	10	675	All	0.692	0.818
10	675	Sub	0.579	0.658	10	675	Sub	0.717	0.840
30	6525	All	0.579	0.662	30	6525	All	0.650	0.767
30	6525	Sub	0.566	0.645	30	6525	Sub	0.657	0.772
90	60075	All	0.550	0.620	90	60075	All	0.601	0.700
90	60075	Sub	0.546	0.616	90	60075	Sub	0.603	0.700
262	512865	All	0.537	0.592	284	602790	All	0.565	0.632
251	470625	Sub	0.536	0.595	271	548775	Sub	0.567	0.637

Appendix 4.13: The pairwise comparison for sectors

2000-Region					**2010-Region**				
Bin	**Pairs**	**Sample**	**Unweighted**	**Weighted**	**Bin**	**Pairs**	**Sample**	**Unweighted**	**Weighted**
2	105	All	0.429	0.430	2	171	All	0.515	0.552
2	105	Sub	0.448	0.446	2	171	Sub	0.544	0.595
4	630	All	0.435	0.419	4	1026	All	0.480	0.486
4	630	Sub	0.467	0.466	4	1026	Sub	0.512	0.521
10	4725	All	0.441	0.403	10	7695	All	0.467	0.438
10	4725	Sub	0.468	0.456	10	7695	Sub	0.500	0.501
30	45675	All	0.452	0.407	30	74385	All	0.476	0.446
30	45675	Sub	0.479	0.466	30	74385	Sub	0.496	0.493
90	420525	All	0.464	0.420	90	684855	All	0.478	0.448
90	420525	Sub	0.484	0.469	90	684855	Sub	0.497	0.497
338	5980065	All	0.479	0.442	337	9681336	All	0.482	0.456
296	4584300	Sub	0.489	0.475	296	7465860	Sub	0.493	0.487
2000-Agglomeration					**2010-Agglomeration**				
Bin	**Pairs**	**Sample**	**Unweighted**	**Weighted**	**Bin**	**Pairs**	**Sample**	**Unweighted**	**Weighted**
2	105	All	0.467	0.462	2	171	All	0.544	0.581
2	105	Sub	0.505	0.551	2	171	Sub	0.579	0.624
4	630	All	0.429	0.410	4	1026	All	0.460	0.480
4	630	Sub	0.470	0.477	4	1026	Sub	0.494	0.515
10	4725	All	0.427	0.395	10	7695	All	0.444	0.444
10	4725	Sub	0.443	0.424	10	7695	Sub	0.452	0.462
30	45675	All	0.438	0.406	30	74385	All	0.458	0.452
30	45675	Sub	0.443	0.416	30	74385	Sub	0.472	0.473
90	420525	All	0.449	0.412	90	684855	All	0.468	0.458
90	420525	Sub	0.453	0.419	90	684855	Sub	0.475	0.474
312	5094180	All	0.467	0.439	316	8510670	All	0.475	0.464
282	4160205	Sub	0.470	0.444	287	7018011	Sub	0.483	0.481
2000-City					**2010-City**				
Bin	**Pairs**	**Sample**	**Unweighted**	**Weighted**	**Bin**	**Pairs**	**Sample**	**Unweighted**	**Weighted**
2	105	All	0.457	0.445	2	171	All	0.579	0.623
2	105	Sub	0.467	0.461	2	171	Sub	0.579	0.613
4	630	All	0.437	0.473	4	1026	All	0.487	0.550
4	630	Sub	0.433	0.479	4	1026	Sub	0.485	0.548
10	4725	All	0.436	0.444	10	7695	All	0.459	0.503
10	4725	Sub	0.445	0.458	10	7695	Sub	0.455	0.505
30	45675	All	0.461	0.459	30	74385	All	0.479	0.516
30	45675	Sub	0.459	0.463	30	74385	Sub	0.476	0.514
90	420525	All	0.468	0.465	90	684855	All	0.488	0.518
90	420525	Sub	0.468	0.468	90	684855	Sub	0.488	0.520
262	3590055	All	0.481	0.479	284	6871806	All	0.491	0.511
251	3294375	Sub	0.482	0.483	271	6256035	Sub	0.492	0.513

Appendix 4.14: The pairwise comparison for occupations

2000-Region					2010-Region				
Bin	**Pairs**	**Sample**	**Unweighted**	**Weighted**	**Bin**	**Pairs**	**Sample**	**Unweighted**	**Weighted**
2	21	All	0.286	0.216	2	21	All	0.524	0.624
2	21	Sub	0.381	0.432	2	21	Sub	0.619	0.726
4	126	All	0.357	0.298	4	126	All	0.421	0.453
4	126	Sub	0.405	0.391	4	126	Sub	0.524	0.578
10	945	All	0.363	0.291	10	945	All	0.410	0.407
10	945	Sub	0.426	0.407	10	945	Sub	0.469	0.502
30	9135	All	0.411	0.326	30	9135	All	0.446	0.426
30	9135	Sub	0.449	0.416	30	9135	Sub	0.466	0.472
90	84105	All	0.447	0.386	90	84105	All	0.456	0.432
90	84105	Sub	0.474	0.450	90	84105	Sub	0.485	0.490
338	1196013	All	0.467	0.425	337	1188936	All	0.463	0.439
296	916860	Sub	0.483	0.460	296	916860	Sub	0.480	0.474
2000-Agglomeration					**2010-Agglomeration**				
Bin	**Pairs**	**Sample**	**Unweighted**	**Weighted**	**Bin**	**Pairs**	**Sample**	**Unweighted**	**Weighted**
2	21	All	0.429	0.574	2	21	All	0.524	0.669
2	21	Sub	0.429	0.574	2	21	Sub	0.667	0.793
4	126	All	0.278	0.316	4	126	All	0.381	0.511
4	126	Sub	0.357	0.450	4	126	Sub	0.460	0.583
10	945	All	0.338	0.303	10	945	All	0.419	0.502
10	945	Sub	0.402	0.415	10	945	Sub	0.430	0.507
30	9135	All	0.395	0.353	30	9135	All	0.456	0.507
30	9135	Sub	0.416	0.410	30	9135	Sub	0.460	0.511
90	84105	All	0.434	0.412	90	84105	All	0.465	0.489
90	84105	Sub	0.448	0.441	90	84105	Sub	0.474	0.502
312	1018836	All	0.462	0.444	316	1045170	All	0.480	0.488
282	832041	Sub	0.471	0.467	287	861861	Sub	0.490	0.508
2000-City					**2010-City**				
Bin	**Pairs**	**Sample**	**Unweighted**	**Weighted**	**Bin**	**Pairs**	**Sample**	**Unweighted**	**Weighted**
2	21	All	0.571	0.726	2	21	All	0.619	0.791
2	21	Sub	0.571	0.726	2	21	Sub	0.619	0.791
4	126	All	0.484	0.588	4	126	All	0.484	0.684
4	126	Sub	0.476	0.591	4	126	Sub	0.532	0.728
10	945	All	0.443	0.476	10	945	All	0.453	0.601
10	945	Sub	0.449	0.498	10	945	Sub	0.458	0.611
30	9135	All	0.460	0.481	30	9135	All	0.478	0.583
30	9135	Sub	0.471	0.511	30	9135	Sub	0.486	0.606
90	84105	All	0.475	0.496	90	84105	All	0.499	0.571
90	84105	Sub	0.476	0.505	90	84105	Sub	0.502	0.591
262	718011	All	0.491	0.507	284	843906	All	0.504	0.551
251	658875	Sub	0.492	0.514	271	768285	Sub	0.505	0.558

Chapter 5 Conclusion

Since the end of the cold war, we have seen China as a rising star in terms of economic performance. Expanding Chinese cities have been attracting laborers from rural areas. In addition, workers flow from less developed urban areas to bigger cities in search of career opportunities. High skilled workers are gathering in large cities in which skill intensive sectors and occupations are clustered. The fast urbanization and rapid economic growth were mutually reinforcing, shaping the Chinese economy. In addition, China has been deeply integrated into the world economy and building up a strengthened link with Africa in the waves of globalization. By 2010, China has overtaken European countries and become the second largest export market for Africa, ranking second only after the USA. The profound engagement of China into Africa was not only driven by China's fast growth, but also propelled by China's governmental preferential trade policies. Even though the African exports have achieved remarkable increase, there was no evidence of industry upgrading changes in the African export structure. In particular, most African countries still rely on primary goods of agricultural products or natural resources as their main exports.

5.1 Main findings

We analyze two particular cases of development in this dissertation, namely China and Africa. We focus on three specific topics: (i) China-Africa trade, (ii) the African export structure, and (iii) China's urbanization. This dissertation thus expands our knowledge in three different fields of economics: international trade, development economics, and urban economics.

In chapter 2, we investigate the trade relationships between China and African in the era of globalization, especially linking China's fast growth to its corresponding rise in African exports. On the robust basis of the Gravity Model, we adopt an out-of-sample estimation strategy and compare the predicted export values with actual data. We find that the exceptional economic growth of China alone cannot fully explain the increase of Africa-

China trade. The large gap between the predicted values and the actual values indicates hidden factors having non-negligible impacts on the African exports to China. We summarize this phenomenon as the Africa-China trade puzzle. To better explain the puzzle, we include China's trade policy variables, such as China's outward *FDI* in Africa, setting up *Economic and Trade Cooperation Zones* (ETCZ), *targeted trade liberalization* (duty-free lists), China's *business and politics separation policy* and the *resource-seeking* motive into our analysis. Generally, we find all of the above trade policies are positively related to African exports to China, though the effect of setting up the duty-free list is not significant since only a limited number of African counties and import sectors were included in the duty-free list in our sample period. Overall, the fast economic growth of China cannot fully explain the export trade from Africa to China. The supplement of the basic gravity model with trade policies enhances our understanding of the Africa-China trade puzzle.

In chapter 3, we examine the export structure of Africa and seek to identify potential industrial upgrading types of structural change. Unlike most previous studies that rely on partial information of trade patterns, we use a new method based on the entire distribution of revealed comparative advantages. Consequently, though economic theory suggests industrial structure upgrading often occurs when the economy grows rapidly, we find this is not necessarily the case, at least in Africa. Compared to the benchmarks of OECD, EU-12 and G5 countries, Africa shows modest structural change in its export from 1990 to 2010. After eliminating some data issues and pseudo structural changes only two countries (Uganda and Mauritius) show plausible structural change patterns in their exports. However, there is still no industrial upgrading of structural change in these two countries when we analyze the history of their economic growth and export expansion. Overall, we find no solid evidence for structural change in African exports. This conclusion indicates that there is still a long way for Africa to proceed in the future to switch to an industry dominant structural continent. At the end of chapter 3, we analyze possible reasons for the absence of structural changes in Africa. We believe that the abundant endowments of natural resources and the absence of skilled labor force, certain geographic and institutional factors prevent African firms from moving to manufacture-based exports.

In chapter 4, we empirically analyze the urban development in Chinese cities during the last decade. Based on the theoretical framework of Davis and Dingel (2013), we analyze the distribution of skills, sectors and occupations across Chinese cities. To identify a clear dynamic change in the development of the cluster of laborers and sectors across Chinese cities, we focus on the years 2000 and 2010. First, we find that the cities' skill distribution systematically varies with city size in both years. In particular, larger cities are relatively more skill abundant, in line with the theoretical prediction in Davis and Dingel (2013) and the empirical analysis of the USA. On the contrary, we find evidence of a dynamic evolution in the sectoral and occupational distributions from 2000 to 2010. The distributions of sectors and occupations across Chinese cities confirms the prediction that lager cities specialize more in skill-intensive production in 2010 only, while we do not find a clear systematic variation between the distribution of sectors (occupations) and city size in 2000. In the end, we attribute this phenomenon to the fast urbanization in China accompanied by the development of market economies during this period. More specifically, the fast urbanization in China enables a large number of farmers who used to work in agricultural sectors (occupations) to switch to manufacturing sectors (occupations), especially in large cities, reshaping the distribution of sectors and occupations across cities.

5.2 policy implications

My dissertation has straightforward policy implications. First of all, the positive relationship between trade policies and the increased trade indicates the success of preferential trade policies in promoting international trade. Even though governmental intervention is only relevant in case of market dysfunction in the classic economic theory, we find certain trade policies could be effective. In the near future, trade policies may be applicable to other less developed regions for their integration into the global economy.

Second, economic growth per se does not guarantee industry-upgrading changes in the trade structure, as we have seen in the case of Africa. Barriers such as resource curse, poor institutions and so forth remind us that blind faith in growth figures cannot achieve long run balanced growth. At certain stages of development, we should pay more attention to the structure of the economy instead of the size of the economy. In particular, Africa needs to make an effort to improve its institution quality, education of laborers, and more equalized distribution of resource rents, which are consistent with the policy implications in Wood and Berge (1997) and Wood and Mayer (2001).

Finally, urbanization and a market-oriented economy reallocated the distributions of labor, sectors, and occupations, raising economic growth. Therefore, we should further reform the economic institutions and promote a market economy. A well-functioning market economy is a precondition for the efficient allocation of laborers, sectors, and occupations across cities, which therefore facilitates urbanization and economic growth.

5.3 Limitations and future research

My research inevitably has some caveats. A good understanding of these limitations shed light for the future research. We will discuss limitations from identification strategies, econometric methods and data availability.

First, it is difficult to establish a causal interpretation on the impact of trade policy on China-Africa trade. Even though we find trade policies promote the Africa-China trade connection, we cannot accurately evaluate the impact of trade policies on the rise of African exports since trade policies are determined endogenously and such analysis is prone to the problem of reverse causality. However, it is safe to claim that trade policies are positively related to African exports to China. A common method to cope with reverse causality is to use instrumental variables. However, it is difficulty to find suitable instruments for trade policies. Future research could search for instruments based on geographic or other exogenous information.

Second, the absent data of aid and FDI restricts the study of Africa-China trade. More specifically, aid is often analyzed in the trade literature as it is usually related to a close connection between two countries. Therefore, aid to Africa is expected to promote Africa-China trade flows. However, the data on aid is either unavailable or incomplete and inaccurate, especially for aid from China. Official aid data have been kept as a secret and aid from NGOs and foundations were not accurately compiled. This data issue restricts our analysis of the relationship between trade policy and African exports. Unfortunately we have to leave aid connections out of our analysis. In addition, the lack of data for Chinese outward FDI to Africa between 2001 and 2002 limited our sample and restricted the explanatory power of tests. If possible, we would like to include extended FDI data and information on aid flows into our analysis of Africa-China trade flows in the future.

Third, we need appropriate benchmarks since Markov chains and mobility indices have no critical values. In practice, we adopt developed economies such as G-12 and G-5 as benchmarks, as there are no prior studies using the same methods analyzing developing countries. In the future, we would like to take developing economies and emerging markets as benchmarks, as these share more similarities with Africa. A comparison between Africa and other developing countries that experienced both economic growth and structural change would illustrate the roots of absence of structural changes in Africa in a more compelling way and therefore provide better possible policy implications.

Fourth, and finally, the quality of Chinese data could be improved. The quality of the population census data in 2000 is believed to be low for various reasons. The census of population is conducted every ten years. The rural-urban migration enlarged substantially from a very low level in 1990 to a high level in 2000. The National Bureau of Statistics was blamed to be unprepared for the sampling method for the surging rural-urban migrants.[73] As a consequence, a considerable fraction of rural-urban migrants were omitted in the census in 2000, which is expected to make the data inaccurate. The situation improved in 2010 as the National Bureau of Statistics polished the sampling and other statistical methods. The sampling method in 2010 was nonetheless still problematic in the sense that the data is collected from population residing in a given city for at least five consecutive years. This criterion omits the population that temporarily lives and works in the city. Therefore, accessible micro level data in the future would enable us to perform a better analysis, as this omitted population can then be taken into account.

At this point I conclude my dissertation. I have indicated that there are still many open questions waiting for an answer, which requires time, effort, and facing various challenges. In short, my work is not yet done and I hope to continue this analysis in the future.

[73] See “Make the data in population census more accurate” New York Times.
http://cn.nytimes.com/opinion/20130306/cc06population/

Nederlandse Samenvatting (Dutch Summary)

Mijn dissertatie analyseert economische ontwikkelingen in China en Afrika over de afgelopen 20 jaar. De sterke groei van de onderlinge handel heeft er voor gezorgd dat China nu de belangrijkste handelspartner is van Afrika, mede dankzij de handelsbevorderende maatregelen. China kan in dit opzicht een voorbeeld zijn voor andere landen betreffende integratie in het wereldhandelssysteem. We analyseren niet alleen de omvang van de Afrikaanse handel maar ook de structuur daarvan, aangezien verbetering van de industriële structuur van de export cruciaal is voor continuering van de groei. We vinden, helaas, dat deze industriële verandering in Afrika nog niet plaats vindt. Mogelijke oorzaken zijn (i) de beperkte beschikbaarheid van goed geschoolde arbeid, (ii) bepaalde institutionele en geografische factoren en (iii) de 'vloek' van overvloeidig aanwezige grondstoffen. Tenslotte kijken we naar de evolutie van de agglomeratie van economische activiteiten in China, die minder geconcentreerd is dan die van andere industriële landen in een soortgelijke fase van ontwikkeling. Dit wordt vaak toegeschreven aan institutionele factoren, met name het hukou systeem. Ons onderzoek toont echter aan dat de grotere steden in China relatief overvloediger geworden zijn in hooggeschoolde arbeid van 2000 tot 2010, met de daarbij behorende beroepen en produktie in sectoren die intensief gebruik maken van die hooggeschoolde arbeid. Wij zien dit als indicatie dat de chinese economie meer markt-georiënteerd wordt over de tijd en dat de agglomeratievoordelen gaandeweg meer tot hun recht komen.

Curriculum Vitae

Shiwei Hu (1985) was born in Fushun, China. She had undergraduate studies in Agriculture at China Agricultural University in Beijing and received her bachelor degree in 2008. From 2008 to 2010, she completed her master education in Economics at China Agricultural University. Over the period 2010-2014, Shiwei was a PhD candidate at Utrecht University School of Economics at the Chair of international macroeconomics, where her PhD study was funded by the China Scholarship Council (CSC). During her PhD, Shiwei Hu visited Xi'an Jiaotong-Liverpool University in China as a visiting researcher for a month. Her research interests cover the international trade of China and Africa, development economics and urban economics of China.

TKI Dissertation Series

USE 001 **Bastian Westbrock** (2010): *Inter-firm networks: economic and sociological perspectives***.**

USE 002 **Yi Zhang** (2011): *Institutions and International Investments: Evidence from China and Other Emerging Markets*.

USE 003 **Ryan van Lamoen** (2011): *The Relationship between Competition and Innovation: Measuring Innovation and Causality.*

USE 004 **Martijn Dröes** (2011): *House Price Uncertainty in the Dutch Owner-Occupied Housing Market.*

USE 005 **Thomas van Huizen** (2012): *Behavioural Assumptions in Labour Economics: Analysing Social Security Reforms and Labour Market Transitions.*

USE 006 **Martijn Boermans** (2012): *International Entrepreneurship and Enterprise Development.*

USE 007 **Joras Ferwerda** (2012): *The Multidisciplinary Economics of Money Laundering.*

USE 008 **Federico D'Onofrio** (2013): *Observing the country: a history of Italian agricultural economics, 1900-1930.*

USE 009 **Saraï Sapulete** (2013): *Works Council Effectiveness: Determinants and Outcomes.*

USE 010 **Britta Hoyer** (2013): *Network Formation under the Threat of Disruption.*

USE 011 **Coen Rigtering** (2013): *Entrepreneurial Orientation: Multilevel Analysis and Consequences.*

USE 012 **Beate Cesinger** (2013): *Context and Complexity of International Entrepreneurship as a Field of Research.*

USE 013 **Jan de Dreu** (2013): *Empirical essays on the governance of financial institutions.*

USE 014 **Lu Zhang** (2013): *Industrial Specialization: Determinants, Processes and Consequences.*

USE 015 **Matthias Filser** (2013): *Strategic Issues in Entrepreneurship and Family Business Research.*

USE 016 **Mikko Pohjola** (2013): *A Compilation of Studies on Innovation in Firms: Capabilities, Strategies, and Performance*.

USE 017 **Han-Hsin Chang** (2013): *Heterogeneity in Development.*

USE 018 **Suzanne Heijnen** (2014): *Analyses of sickness absence.*

USE 019 **Mark Kattenberg** (2014): *The Economics of Social Housing: Implications for Welfare, Consumption, and Labor Market Composition.*

USE 020 **Daniel Possenriede** (2014): *The Economics of Temporal and Locational Flexibility of Work.*

USE 021 **Dirk Gerritsen** (2014): *The Relevance of Security Analyst Opinions for Investment Decisions.*