Three essays on human mobility and child development in China

Huaxin Wanglu

http://hdl.handle.net/10803/688965

Data de defensa: 26-07-2023

UNIVERSITAT RAMON LLULL

DOCTORAL THESIS

Title	Three Essays on Human Mobility and Child Development in China
Presented by	Huaxin Wanglu
Centre	IQS School of Management
Department	Department of Quantitative Methods
Directed by	Prof. Octasiano Miguel VALERIO MENDOZA, Prof. Flavio COMIM

Acknowledgements

I am deeply indebted to Prof. Octasiano Miguel VALERIO MENDOZA, my primary advisor at Institut Químic de Sarrià of Universitat Ramon Llull, for his guidance, supervision, and multifarious support throughout my Ph.D. journey. I lacked confidence in myself to accomplish this intellectual challenge until he believed in me. The year 2020 was bad timing for Ph.D. applications as I lost several great opportunities due to COVID-19 outbreaks. It is fortunate for me to join the CHINEQUALJUSTIC project in the end and have him as my supervisor. He was always responsive to my questions and offered thoughtful counsel, being not only serious with my research but also my future. Likewise, I am grateful for my subadvisor, Prof. Flavio COMIM, who generously contributed time and ideas to this dissertation and learned about my personal issues in earnest and with patience. I will always remember the dinner I had with him and his family on April 5th, 2023. It was my first time having meals with non-Chinese people in a family atmosphere. I must also thank Prof. Mihály Tamás BORSI, who is not officially my subadvisor but I consider him to be one, for his great help with improving this manuscript and for his invaluable, warmhearted advice and assistance during my job search.

I am also highly appreciative to Prof. Simiao CHEN, from the Heidelberg Institute of Global Health of Heidelberg University, for hosting me and sharing publishing tips; to Prof. Vicente ROYUELA, from the School of Economics of Universitat de Barcelona, for providing constructive feedback and suggestions whereby I largely improved **Chapter 1** of this thesis; to Prof. Lex(Laixun) ZHAO, co-editor of the Journal of International Trade & Economic Development, and the anonymous reviewer for helping me publish the first chapter. A large number of colleagues and session participants at a variety of conferences, workshops, and seminars who delivered questions, feedback, and encouragement have also left a mark on this work.

Last but most importantly, I wish to thank my parents, Wei LU and Huiming WANG, for their love, mental support, and financial aid. In particular, my mother has been my most steadfast comrade who always stands by me, empathizing with my feelings and supporting me in all my decisions. Without her support and encouragement, I won't be able to fulfill this undertaking. I would also like to convey my gratitude to my friends, especially Cencen HUANG, Ziqing LIAO, and Zhifei LIU, who spent countless time chatting and/or hanging out with me, taking me out of the isolation and bringing immeasurable joy to the three years of my life.

Abstract

Human mobility and child development are essential parts of human development. China, as the world's second most populous nation, and currently the second largest and upper-middle-income economy, provides a unique context for researching these two topics against the backdrop of structural changes. This thesis endeavors to advance our understanding of the intricacies behind human mobility and child development in China from three perspectives: labor migration, travel dynamics, and child academic development.

Co-authored with Octasiano M. VALERIO MENDOZA, **Chapter 1** investigates the effects of job prospects on individual migration decisions across prefecture cities. To this end, we created proxy variables for wage and employment prospects, introduced reference dependence to a dynamic discrete choice model, and estimated corresponding empirical specifications with a unique quasi-panel of 66,427 individuals from 283 cities during 1997–2017. To address multilateral resistance to migration resulting from future attractiveness, we exploited various monadic and dyadic fixed effects. Multilevel logit models and two-step system GMM estimation were adopted for the robustness check. Our primary findings are that a 10% increase in the ratio of sector-based employment prospects in cities of destination to cities of origin raises the probability of migration by 1.281–2.185 percentage points, and the effects tend to be stronger when the scale of the ratio is larger. Having a family migration network causes an increase of approximately 6 percentage points in migratory probabilities. Further, labor migrants are more likely to be male, unmarried, younger, or more educated.

Co-authored with Mihály Tamás BORSI and Octasiano M. VALERIO MENDOZA, **Chapter 2** investigates intra-city mobility trajectories of 368 Chinese cities within a non-

linear time-varying latent factor framework during the end of the Zero-COVID Policy to uncover if regional disparities tend to widen and to identify cities that are falling behind. To this end, we compiled a novel panel using the latest Baidu Mobility Data and the risk-level data published by the State Council of the People's Republic of China. Further, it examines the causal effects of exposure to high COVID-19 risk in the city on commuting behaviors. Our main results show that gaps in local travel strength tend to decrease within each cluster but widen between different clusters. Moreover, the "high-risk" alert, which can trigger the implementation of the most stringent containment measures set by the Zero-COVID Policy, had persistently dampened home-workplace commuting rates during the studied sample period. In addition, divisions in intra-city travel strength and commuting rates between Western and the rest of China have been identified. In sum, this chapter suggests that socio-economic activities which depend heavily on human mobility are recovering at different rates across clusters, with less travel-intensive cities, such as Lhasa, Shihezi, and Urümq, increasingly left behind.

Co-authored with Flavio COMIM and Octasiano M. VALERIO MENDOZA, **Chapter 3** provides a composite analysis of children's academic development based on comprehensive outcomes, including valuable processes, and introduces a series of innovative indicators rooted in the capability approach. Significantly, it presents an index of parents' advantages to capture how parents influence their children and establishes a new indicator of spending priorities to reify the value of children's education that families have reasoned. The study sample consists of 8,422 children and adolescents surveyed from 2012 to 2018. Our results show that a 1% increase in the parent advantage index yields an increase of 13.85% to 21.31% in children's academic development, and the biggest leap in prioritizing education-relevant spending increases the child outcomes by 2.88% to 6.57%. Further, if the interaction of parents' higher educational attainment and spending priorities is taken into account, children's academic development could reach a maximum difference of 21.96%. In addition, prioritizing spending on healthcare- and clothing-relevant items also influences educational development. We conclude that parents can influence the development of their children through their beings and doings and, particularly, the value they place on their children's education, which goes beyond a limited focus on material dimensions, such as income and investment levels in education.

In sum, this thesis reveals that China's current industrial reform and COVID-related strategies, including reopening, can influence migratory and non-migratory human mobility at the city level. Further, it demonstrates that promoting the value of education and ensuring that families from different backgrounds understand its importance can help reduce disparities in child academic outcomes.

Resumen

La movilidad humana y el desarrollo infantil son partes esenciales del desarrollo humano. China, como la segunda nación más poblada del mundo y actualmente la segunda economía de ingresos medios y grandes, ofrece un contexto único para investigar estos dos temas en el contexto de los cambios estructurales. Esta tesis se esfuerza por mejorar nuestra comprensión de las complejidades detrás de la movilidad humana y el desarrollo infantil en China desde tres perspectivas: migración laboral, dinámica de viajes y desarrollo académico infantil.

Co-escrito con Octasiano M. VALERIO MENDOZA, **Capítulo 1** investiga los efectos de las perspectivas laborales en las decisiones individuales de migración entre ciudades prefecturas. Para ello, creamos variables proxy para los salarios y las perspectivas laborales, introducimos la dependencia de referencia en un modelo de elección discreta dinámico y estimamos especificaciones empíricas correspondientes con un panel cuasi-único de 66,427 individuos de 283 ciudades durante el periodo 1997–2017. Para abordar la resistencia multilateral a la migración resultante de la atracción futura, explotamos varios efectos fijos monádicos y diádicos. Se adoptaron modelos logit multinivel y estimación del sistema GMM de dos pasos para la verificación de la robustez. Nuestros hallazgos principales son que un aumento del 10% en la proporción de perspectivas laborales basadas en sectores en las ciudades de destino con respecto a las ciudades de origen aumenta la probabilidad de migración en 1.281–2.185 puntos porcentuales, y los efectos tienden a ser más fuertes cuando la escala de la proporción es mayor. Tener una red de migración familiar causa un aumento de aproximadamente 6 puntos porcentuales en las probabilidades migratorias. Además, los migrantes laborales son más propensos a ser hombres, solteros, más jóvenes o más educados.

Co-escrito con Mihály Tamás BORSI y Octasiano M. VALERIO MENDOZA, **Capítulo 2** investiga las trayectorias de movilidad intra-ciudad de 368 ciudades chinas dentro de un marco de factor latente no lineal y variable en el tiempo durante el final de la Política de Cero-COVID para descubrir si las disparidades regionales tienden a ampliarse e identificar las ciudades que están rezagadas. Para ello, compilamos un panel novedoso utilizando los últimos Datos de Movilidad de Baidu y los datos de nivel de riesgo publicados por el Consejo de Estado de la República Popular China. Además, examina los efectos causales de la exposición al alto riesgo de COVID-19 en la ciudad sobre los comportamientos de desplazamiento. Nuestros principales resultados muestran que las brechas en la fortaleza del viaje local tienden a disminuir dentro de cada grupo, pero se amplían entre diferentes grupos. Además, la alerta de "alto riesgo", que puede desencadenar la implementación de las medidas de contención más estrictas establecidas por la Política de Cero-COVID, había disminuido persistentemente las tasas de desplazamiento entre el hogar y el lugar de trabajo durante el período de muestra estudiado. Además, se han identificado divisiones en la fuerza de viaje intra-ciudad y las tasas de desplazamiento entre el oeste y el resto de China. En resumen, este documento sugiere que las actividades socioeconómicas que dependen en gran medida de la movilidad humana se están recuperando a diferentes velocidades en los diferentes grupos, dejando cada vez más atrás a las ciudades con menos intensidad de viaje, como Lhasa, Shihezi y Urümqi.

Co-escrito con Flavio COMIM y Octasiano M. VALERIO MENDOZA, el **Capítulo 3** ofrece un análisis compuesto del desarrollo académico de los niños basado en resultados integrales, incluyendo procesos valiosos, e introduce una serie de indicadores innovadores basados en el enfoque de capacidades. Significativamente, presenta un índice de ventaja de los padres para capturar cómo influyen los padres en sus hijos y establece un nuevo indicador de prioridades de gasto para reificar el valor de la educación de los niños que las familias han razonado. La muestra de estudio consta de 8,422 niños y adolescentes encuestados desde 2012 hasta 2018. Nuestros resultados muestran que un aumento del 1% en el índice de ventaja de los padres produce un aumento del 13.85% al 21.31% en el desarrollo académico de los niños, y el mayor salto en la priorización del gasto relevante para la educación aumenta los resultados de los niños en un 2.88% a un 6.57%. Además, si se tiene en cuenta la interacción entre el logro educativo superior de los padres y las prioridades de gasto, el desarrollo académico de los niños podría alcanzar una diferencia máxima del 21.96%. Además, la priorización del gasto en artículos relacionados con la salud y la ropa también influye en el desarrollo educativo. Concluimos que los padres pueden influir en el desarrollo de sus hijos a través de su ser y hacer, y especialmente, del valor que otorgan a la educación de sus hijos, lo que va más allá de un enfoque limitado en dimensiones materiales, como el ingreso y los niveles de inversión en educación.

En resumen, esta tesis revela que la reforma industrial actual de China y las estrategias relacionadas con COVID, incluida la reapertura, pueden influir en la movilidad humana migratoria y no migratoria a nivel de la ciudad. Además, demuestra que promover el valor de la educación y asegurarse de que las familias de diferentes orígenes comprendan su importancia puede ayudar a reducir las disparidades en los resultados académicos de los niños.

Resum

La mobilitat humana i el desenvolupament infantil són parts essencials del desenvolupament humà. La Xina, com la segona nació més poblada del món i actualment la segona economia més gran i de renda mitjana alta, proporciona un context únic per investigar aquests dos temes en el marc de canvis estructurals. Aquesta tesi es proposa avançar en la comprensió de les complexitats de la mobilitat humana i el desenvolupament infantil a la Xina des de tres perspectives: la migració laboral, la dinàmica dels viatges i el desenvolupament acadèmic infantil.

Escrit conjuntament amb Octasiano M. VALERIO MENDOZA, **Capítol 1** investiga els efectes de les perspectives laborals en les decisions migratòries individuals a través de les ciutats prefectura de la Xina. Per a això, vam crear variables de proxy per a les perspectives salarials i d'ocupació, vam introduir la dependència de referència en un model de selecció discret dinàmic i vam estimar especificacions empíriques corresponents amb un quasi-panell únic de 66,427 individuals de 283 ciutats durant el període 1997–2017. Per abordar la resistència multilateral a la migració que resulta de l'atractiu futur, vam explotar diversos efectes fixes monàdics i diàdics. Per a la verificació de la robustesa, es van adoptar models logit multinivell i l'estimació del sistema GMM de dos passos. Les nostres troballes principals són que un augment del 10% en la relació de les perspectives d'ocupació basades en el sector en les ciutats de destinació respecte a les ciutats d'origen augmenta la probabilitat de migració en 1.281–2.185 punts percentuals, i els efectes tendeixen a ser més forts quan l'escala de la relació és més gran. Tenir una xarxa de migració familiar provoca un augment d'aproximadament 6 punts percentuals en les probabilitats migratòries. A més, els treballadors migrants són més propensos a ser homes, solters, més joves o més instruïts.

Co-autoria amb Mihály Tamás BORSI i Octasiano M. VALERIO MENDOZA, **Capítol 2** investiga les trajectòries de mobilitat intra-ciutat de 368 ciutats xineses en un marc de factor latent no lineal i variable en el temps durant la fi de la Política Zero-COVID per descobrir si les disparitats regionals tenen tendència a augmentar i identificar les ciutats que es queden enrere. Per a això, vam compilar un nou panell utilitzant les últimes dades de mobilitat de Baidu i les dades de nivell de risc publicades pel Consell d'Estat de la República Popular de la Xina. A més, examina els efectes causals de l'exposició a un alt risc de COVID-19 a la ciutat sobre els comportaments de desplaçament. Els nostres resultats principals mostren que les diferències en la força del trànsit local tendeixen a disminuir dins de cada grup, però a augmentar entre diferents grups. A més, l'alerta "alt risc" que pot desencadenar la implementació de les mesures de contenció més estrictes establertes per la Política Zero-COVID, ha frenat persistentment les taxes de desplaçament entre el lloc de treball i el domicili durant el període d'estudi. A més, s'han identificat divisions en la força del trànsit intra-ciutat i les taxes de desplaçament entre el Western i la resta de la Xina. En resum, aquest article suggereix que les activitats socioeconòmiques que depenen força de la mobilitat humana es recuperen a diferents ritmes a través de grups de ciutats, amb ciutats menys intensives en viatges, com Lhasa, Shihezi i Urümq, que es queden cada vegada més enrere.

Coautoritzat amb Flavio COMIM i Octasiano M. VALERIO MENDOZA, el **Capítol 3** proporciona un anàlisi compost del desenvolupament acadèmic dels infants basat en resultats comprehensius, incloent processos valuosos, i introdueix una sèrie d'indicadors innovadors arrelats en l'enfocament de capacitat. D'una manera significativa, presenta un índex d'avantatges dels pares per capturar com els pares influeixen en els seus fills i estableix un nou indicador de prioritats de despesa per reificar el valor de l'educació dels nens que les famílies han raonat. La mostra d'estudi consta de 8,422 nens i adolescents enquestats des de 2012 fins a 2018. Els nostres resultats mostren que un augment del 1% en l'índex d'avantatge dels pares produeix un augment del desenvolupament acadèmic dels nens del 13.85% al 21.31%, i el salt més gran en prioritzar les despeses relacionades amb l'educació augmenta els resultats dels nens del 2.88% al 6.57%. A més, si es té en compte la interacció entre l'educació superior dels pares i les prioritats de despesa, el desenvolupament acadèmic dels nens podria arribar a una diferència màxima del 21.96%. A més, prioritzar la despesa en articles relacionats amb la salut i la roba també influencia el desenvolupament educatiu. Concloem que els pares poden influir en el desenvolupament dels seus fills a través del seu ser i fer, i particularment pel valor que atorguen a l'educació dels seus fills, que va més enllà d'un enfocament limitat en les dimensions materials, com ara els nivells d'ingressos i inversió en l'educació.

En resum, aquesta tesi revela que la reforma industrial actual de Xina i les estratègies relacionades amb la COVID, incloent la reobertura, poden influir en la mobilitat humana migratòria i no migratòria a nivell de ciutat. A més, demostra que promoure el valor de l'educació i assegurar-se que les famílies de diferents orígens comprenguin la seva importància pot ajudar a reduir les disparitats en els resultats acadèmics dels nens.

Contents

List of Tables

List of Figures

Introduction

Human mobility and child development are engines of human development and economic growth (e.g., De Haas 2009; Young et al. 2002). On one hand, the movement of human beings between spatial units is an important mechanism through which individuals gain access to social and material resources, unlock new opportunities, and improve their quality of life. On the other hand, childhood and adolescence are the most malleable stages of human development. Valuable capabilities developed during these early periods can pave the way to flourishing adulthood, including access to higher education and a rise in lifetime earnings.

Notably, human mobility is often studied as international migration. Yet for citizens in many developing countries whose passports are not as powerful as those of developed nations, international migration, which predominates in migration research (United Nations Development Programme 2020), is not a real opportunity for the majority.[1] Instead, internal migration is an option much more obtainable to them. In particular, for large countries like China, moving across cities can be as distant as an international journey elsewhere, and socio-economic differences, including job opportunities, living standards, and urban amenities, between sending and receiving locations can be equally remarkable. Thus, internal migration especially within developing countries is pressing but relatively understudied.

Further, non-migratory travel, such as commuting, which is more closely related to people's daily lives, is also an intrinsic component of human mobility. It has been heavily emphasized as a primary means to prevent transmission since the outbreak of COVID-19. However, disruptions to travel are also detrimental to both individual

[1] For instance, searching "international migration" on Scopus returns 8,183 articles in Social Sciences, tripling the amount of literature on "internal (domestic) migration".

well-being and economic activities (e.g., Krauss et al. 2022; Sunio et al. 2023; Wu et al. 2023), suggesting closer attention to mobility recovery. While China just terminated its Zero-COVID Policy on December 07, 2022, providing a unique scenario worth scrutiny.

In addition, education, one of the three dimensions of the Human Development Index (HDI), lays the groundwork for children's development. Multi-dimensional poverty can hardly be addressed without equalizing educational opportunities. Chinese society also provides an interesting context in this regard, in terms of its exceptional emphasis on the national college entrance examination, the dominant role of educational attainments in its labor markets, and the Double Reduction Policy announced on July 24, 2021, which includes a ban on for-profit after-school tutoring on weekends, public holidays and school breaks, aimed at alleviating children's workloads and parents' financial burdens.

In sum, this thesis dives into issues on human mobility and child development in China from three perspectives: labor migration, local travel dynamics, and academic development of children and adolescents. Firstly, economic reasons, typically income gaps, are found to be the main drivers of migration decisions. However, as Chinese society sees more people moving beyond the poverty trap and attaining higher levels of education than their previous generations, the driving forces behind their migration decisions, even within the scope of economic motives, can be quite different from those in the past. As long as China continues to make efforts to eliminate poverty and advance its human capital growth, the observed trend suggests that the decision-making of more educated and wealthier migrants who care about their future utilities could be more strategic and visionary than the myopic assumption generally postulated in migration literature (Beine et al. 2016). However, the role of individual expectations in labor migration decisions has not yet been documented in Chinese research until this thesis. Further, rural-to-urban migration is the most widely studied type of migration in Chinese literature (e.g., Cheng et al. 2014; Ye et al. 2016) due to its significant contribution to China's economic development and urbanization. Empirical studies, at the same time, often define internal migrants in China as those who move across provinces (Su et al. 2018). However, with the urban population steadily increasing (the urbanization rate of permanent residents was 64.72% in 2021) and the leading role of prefecture cities in managing Hukou registration, the scarcity of research on urban-to-urban migration or a combination of two types, particularly at the city level, calls for further studies. These dearths and gaps in literature give rise to **Chapter 1**, which is devoted to investigating the influence of job prospects on migration decisions of both rural and urban individuals who move across prefecture boundaries. In particular, the study incorporates reference dependence into a dynamic discrete choice model and formulates empirical specifications with theoretical micro-foundations. To conduct the analysis, it constructs a quasi-panel using the 2017 China Household Finance Survey (CHFS) and longitudinal statistics spanning 1997–2017 from 283 cities and adopts econometric techniques including fixed-effects estimators, multilevel logistic regression, and GMM estimation. In addition, it also takes into account migration networks, which are major determinants of migration beyond economic opportunities, in two ways. Firstly, it exploits various structures of dyadic fixed effects to capture historic migration relationships between origin and destination as well as time-variant heterogeneity in origin (destination) networks. Secondly, considering the salient role of kinship in shaping social networks in China (Peng 2004), family migration networks are measured for each individual to incorporate both the direct effects of migrant relatives and the indirect effects of social networks that individuals can access through their family members.

Secondly, population movement is a major medium for the emergence of human activities. The existing COVID-19 literature on it is concentrated on its relationships with socio-economic characteristics, travel restrictions including stay-at-home orders and people's compliance with them, COVID-19 spread and health outcomes including confirmed cases, deaths, and vaccination, CO_2 emissions, and work-from-home during containment stages. To date, travel behaviors during the transitional period, which can shed light on the recovery of various aspects of individual lives from the pandemic, have been extremely overlooked even on a global scale and is non-existent for China. Further, the risk-level system, the major component of the Zero-COVID Policy which can trigger the introduction of restrictive measures to specific areas, is almost never mentioned in Chinese COVID-19 research. To account for the discussed lack of research, **Chapter 2** is set up to accomplish two sub-objectives: a) assess human mobility trajectories of 368 Chinese cities around the end of the Zero-COVID Policy (January 17, 2022–March 12, 2023) and b) examine the causal effects of exposure to high COVID-19 risk in the city on local travel behaviors during the strict periods (May 17, 2021–June 26, 2022). For these purposes, a sophisticated clustering technique proposed by Phillips and Sul (2007, 2009) and Callaway and Sant'Anna (2021) different-in-difference (DiD) with multiple periods estimator are performed for empirical analyses. The mobility and risk-level data used are retrieved from Baidu Qianxi and the State Council of the People's Republic of China (PRC), respectively.

Thirdly, the bulk of previous studies on child development find that family income and investment made in education grant (dis)advantages to child outcomes, such as schooling decisions and cognitive skills. With this consideration, the Chinese government has recently banned private after-school tutoring to alleviate the impacts of family wealth disparities on children's academic performance. While the resources parents own and the materials and efforts they eventually put into cultivating their children are not always harmonized. In other words, the extent to which family resources, including material and mental support, can be converted into children's functioning and capabilities largely depends on the value parents place on their children's education. However, as is difficult to measure, empirical research on how families value education remains vacant. This is at the heart of **Chapter 3** where the prioritization of household spending on education relative to other items, and parent advantages, modeled on the HDI, are first created to examine how parents' beings and doings, particularly the value of education they have reasoned, influence the development of their children. The study bases its conceptual framework on the Capability Approach and develops several innovative indicators accordingly to reflect Amartya Sen's emphasis on informational pluralism and distinction between culmination and comprehensive outcomes. It employs fixed-effects estimators, multilevel modeling, and Lewbel's (2012) heteroskedasticity-based IVs with longitudinal individual and household data obtained from the China Family Panel Studies (CFPS).

In general, I focus this thesis on quantitative research using nationally-representative micro data and longitudinal city statistics to reinforce causal inference and the generalizability of the findings. Insights drawn from each chapter can be informative for policy-making in China and potentially in other developing countries. I summarize them in order: a) focusing resources on accelerating employment in specific sectors could be beneficial for economically less-developed cities to attract labors; b) tailored travel policies can be established for identified clusters to meet the specific needs of each cluster, and more supportive policies, such as discounts for local transportation and subsidies for vulnerable populations, could be introduced right after the loosening of strict containment measures to make up for the loss of mobility and consumption, and in a progressive manner to balance public health and economic vitality; c) in addition to equalizing educational expenditures, subsequent policies to highlight the value of education among individuals and households that disregard it are helpful.

Chapter 1

Job Prospects and Labor Mobility in China

1.1 Introduction

Countries often undergo a drastic change in their employment structures at times of industrial transformation, entailing labor redistribution. China is a remarkable case for study under this background. In terms of China Statistical Yearbooks, since the beginning of its reforms and opening-up in late 1978, the GDP-based ratio of the primary, secondary and tertiary sector changed from 31:47:22 in 1979 to 8:38:54 in 2020. This indicates that the tertiary sector almost trebled its contribution to national GDP during this period. In recent years, after being the "world's factory" for decades, the central government launched the "Made in China 2025" and "Dual Circulation" strategies to promote upgrading from a labor-intensive, export-oriented manufacturing economy to a service and consumption-driven one.

At the same time, China unveiled a proposed revision to the law on vocational education, announcing in 2020 the "Vocational Education Quality Improvement Action Plan" to fill gaps in skilled technicians and to differentiate skillsets across college graduates. On the demand side, regions at different levels of development require different types of skills at different levels of demand. For instance, inland provinces need workers with plant-based skills in response to the relocation of numerous factories from eastern areas taking advantage of lower labor costs (Qu et al. 2013), whereas the Yangtze River Delta is dedicated to attracting high-tech and managerial professionals for highly-developed manufacturing and service industries (Wang et al. 2020). On the supply side, the job prospects for individuals with diverse skills and profiles differ across regions.

In general, the employment situation of workers with low and average skills is dim, because of the widespread use of automated technology, although the service sector is creating new low-tech jobs (Li et al. 2020). More recently, computerization has added to jobless growth.[1] Frey and Osborne (2017) assert that the majority of routine jobs in manufacturing and a range of sub-sectors within the tertiary sector, such as finance, logistics, and administrative support, are at high risk of being replaced by artificial intelligence (AI) technology, a potential concern for China. Zhou et al. (2020) even suggest that AI technology will replace approximately 278 million jobs in China by 2049. While the creation of non-routine jobs was stagnant during 1990 to 2015, with more than 50% of employment routinized in 2015 (Ge et al. 2021).[2]

The extent of all these (expected) impacts varies across sectors and regions. It is inevitable that employment prospects will be subject to changes in regional employ-

[1] Jobless growth means that an economy is growing at a reasonable rate without the proportionate creation of new jobs.

[2] Tasks that rely on well-defined procedures and activities are classified as routine while tasks that require creativity, problem-solving or human interaction are classified as non-routine. Both tasks further subdivide into cognitive or manual skill types.

ment structures, stimulating relocation, a plausible reaction to worsening prospects in particular regions. These impacts could go beyond concerns about job opportunities into job mobility and/or security. In other words, labor market conditions tend to be inconsistent with economic growth, at least in the short term (e.g., Şahin et al. 2015). These come to our core idea that an employment flourishing industry could provide better job prospects for individuals with migration intentions over those whose job growth shrinks or stagnates, *ceteris paribus*. As seen in Figure 1.1, GDP and wages grew smoothly in Beijing, Shanghai, Guangzhou, and Shenzhen, while employment levels in the four cities fluctuated. While on the other hand, cities with more increases in wages could have a "comparative advantage" to further benefit job prospects of their workers. Thus, we also consider expected earnings as a useful complementarity to the assessment of job prospects of potential migrants. In sum, we divide job prospects into two components, that is, employment and wage prospects.

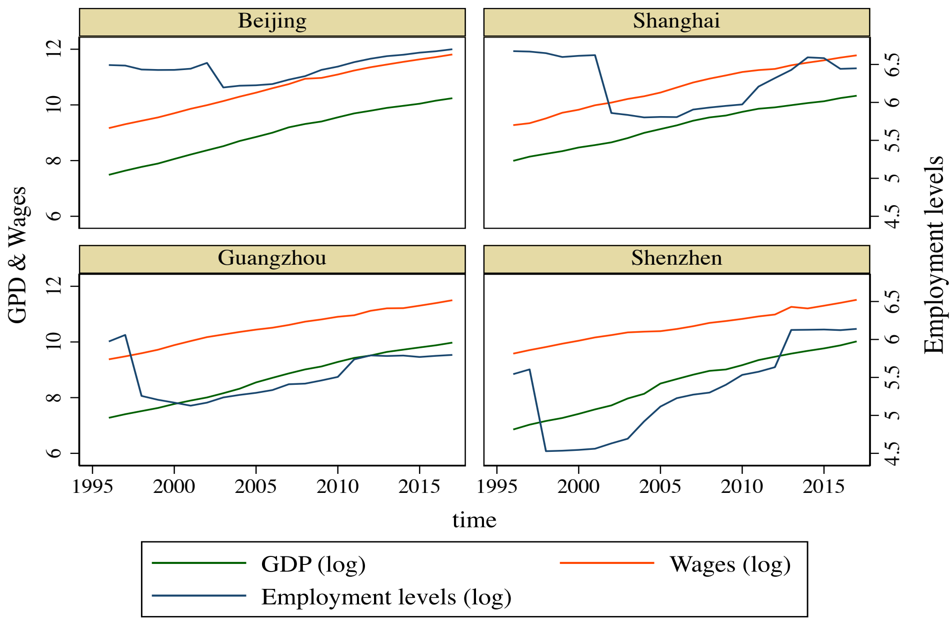


Figure 1.1: GDP, wages vs. employment levels in "Bei-Shang-Guang-Shen".

Source: Author's elaboration using China Data Institute (2021).

Then, the underlying question is how potential migrants make their assessments. For simplicity, we use income as an example for elaboration. Suppose that all individuals in city A always earn 2000 Yuan per month. It will be difficult for them to make an assessment of economic conditions, assuming no inflation. At the same time, people in city B earn 2000 Yuan per month in 2011 but 1500 Yuan in 2012. The latter population is very likely to believe that the economy is getting worse, at least, more likely than are A residents. As prospect theory reveals (Kahneman and Tversky 1979),[3] the deviation $(1500 - 2000)$ from the 2011 income, a reference point closest to the 2012 income, generates a negative signal for evaluation and accordingly incurs a pain of loss to B inhabitants.

Let's consider another two cities C and D where people earn, respectively, 2000 and 3000 Yuan per month in 2011 but both 2500 Yuan in 2012. If B inhabitants with migration intentions are aware of the changes, although moving to either city C or D will lead to the same increases in their income, they will perceive additional gains from C's positive deviation $(2500 - 2000)$. This is because such an uptrend implies that the prospect in city C is better than in the others. Kahneman and Tversky highlighted that individuals would even reverse their preferences when identical outcomes are rephrased as gains or losses. As outlined in Figure 1.2, this example manifests why reference dependence is meaningful in our context.

On the other hand, the feature of population redistribution has also been changing. Intra-provincial migration seems to have been more important than inter-provincial migration for both temporary and permanent migrants during this decade (Meng 2020; Zhang and Zhao 2013). Inter-provincial inequalities have often been associated with the

[3] reference dependence is a central principle in prospect theory. It holds that people evaluate information gradually as losses and gains relative to certain reference points or a status quo, rather than as a final state of an absolute outcome.

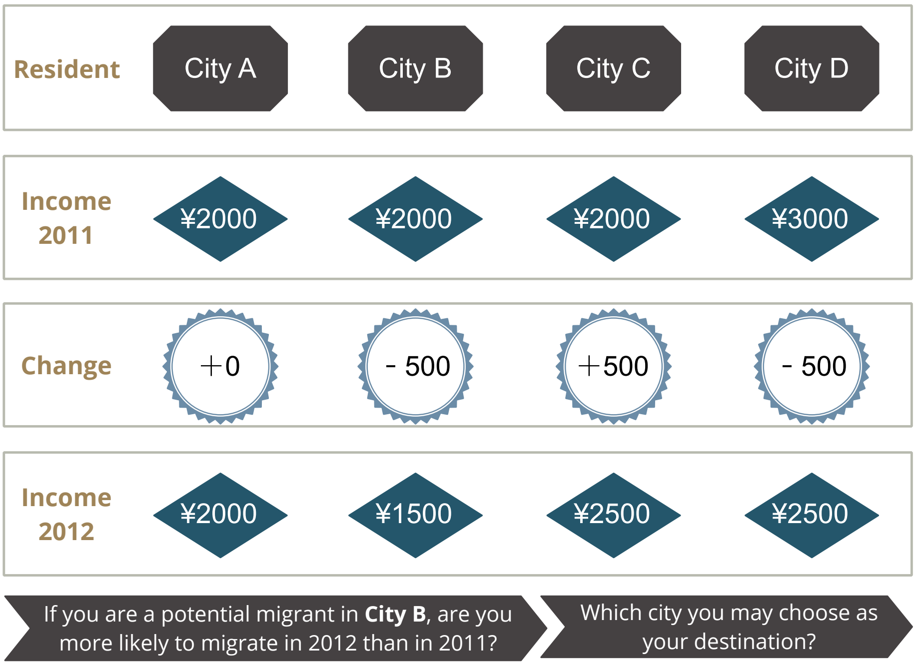


Figure 1.2: Example for understanding reference dependence.

prevalence of inter-provincial migration (e.g., Peng and Swider 2017). If this is the case, a co-movement of the popularity of intra-provincial migration and intra-provincial inequalities will arise. Hence, keeping a watchful eye on city-level migration is of greater importance to China than ever before.[4]

The objective of this chapter is to examine the effects of job prospects on individual migration decisions across prefecture boundaries.[5] Our contribution to the migration debate is fourfold. Firstly, we contributed to the considerable dearth in Chinese migration literature of visionary migration decisions concerned with optimizing future outcomes.[6] Though a handful of studies touch upon migrants' expectations, such as expected land reallocation (e.g., Ren et al. 2020), none considers multilateral resistance

[4]The grant of Hukou is associated with access to a variety of social programs provided by the regional government, such as the entitlement to undertake the college entrance examination (Gaokao), to social security and even to house purchase. Prefecture-level cities are usually the main administrative unit in China designated to manage household registration (Hukou).

[5]In China, prefecture cities rank below provinces, the highest non-national level administrative unit, and above counties.

[6]Migration decisions involving future outcomes across location choices are defined as non-myopic migration (equivalent to "visionary migration" in this study) in a substantial body of migration literature (e.g., Baldwin 2001; Bertoli et al. 2016). In contrast, "Myopia" is a term referring to migration decisions that only depend on the past and/or current situation.

to migration resulting from the future attractiveness (Bertoli and Fernández-Huertas Moraga 2013).[7] We understand that this is the first paper to illustrate migration decision-making of visionary labor migrants within China. Secondly, we extended the random utility maximization (RUM) model of migration by synthesizing the virtues of dynamic discrete choice modeling framework and reference dependence to further illustrate the formation of individual expectations.[8] To our knowledge, this is also the first time that linkages have been established between the RUM model of migration and prospect theory (Kahneman and Tversky 1979). Thirdly, we derived empirical specifications with theoretical micro-foundations. Lastly, we compiled a unique quasi-panel of 66,427 individuals moving from 283 cities to 279 cities during 1997–2017 and thus combined city-level bilateral variations with individual and household characteristics, a level of analysis not yet undertaken by existing Chinese migration literature.[9] While both the new migration patterns and the essential role of prefecture cities in managing Hukou registration signify the importance of understanding city-level push and pull factors.

1.2 Migration and Expectations

1.2.1 Migration in China

Despite the surge of temporary migration since the late 1980s, this phenomenon was not studied empirically until much later, principally due to the lack of data. The earlier migration literature relied mainly on macro-data, typically the national popula-

[7] The term "multilateral resistance to migration" is used to explain influences exerted by the attractiveness of alternative locations on migration rates between any pair of regions.

[8] It is also worth noting that the model we developed can be applied to other spatial dimensions, not limited to internal migration or China.

[9] Cross-sectional and/or provincial analyses predominate Chinese literature on migration (Su et al. 2018).

tion census or the 1% population sample survey, treating intra-provincial migrants as "non-movers" (e.g., Poncet 2006). Although more micro-data became available in the late 2000s, e.g., the Longitudinal Survey on Rural-Urban Migration in China (RUMiC) initiated in 2006, and the China Migrants Dynamic Survey (CMDS) begun in 2009, the vast majority of data options remain incompatible with bilateral or inter-city studies.[10] More recently, micro-data have been popular for migration analysis as they can help alleviate reverse causality problems and allow researchers to control for individual heterogeneity of relevance. However, longitudinal micro-data are much harder to collect, resulting in only a few migration studies on individual decisions over time. Further, the popularity of the province-level research also stems from the remarkable increase in inter-provincial migrant populations since 1987 (Liang 2001), the predominant group until 2010, which was then outpaced by intra-provincial migration, as seen in Figure 1.3. Similar conclusions can be drawn from national statistics on the rural-to-urban subpopulation, the main focus of Chinese migration studies. By combining the 1990, 2000, and 2010 Chinese Censuses with the 1995, 2005, and 2015 1% population sample surveys, Su et al. (2018) show that intra-provincial rural migration flows have outnumbered inter-provincial flows since 2011 and empirically identify that the likelihood of moving within the province of origin increases for rural migrants who are older, more educated, female, single, or from poorer areas.

Determinants of migration are often conceptualized as push-and-pull factors associated with the sending and receiving locations, respectively. Economic reasons, such as income gaps, job opportunities, and land tenure insecurity, are often the most de-

[10] The location information is usually provided at the province level, meaning that the sending and receiving cities cannot be identified. Further, although investigators provided city information, most often only cities of destination are known to applied researchers, resulting in that the majority of migration studies focus on analyzing the receiving context. Until recently, we have still known less about the reasons that migrants leave their hometowns than why they move to their areas of destination.

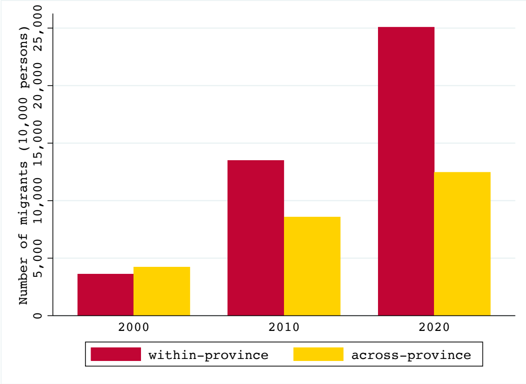


Figure 1.3: The number of intra-provincial migrants vs. inter-provincial migrants.

Source: The 2000, 2010 and 2020 Chinese Censuses.

cisive factors in a substantial body of literature (e.g., Liu and Shen 2014). However, even though migrants usually move for monetary reasons, contributing factors vary between groups and over time. Demographically, researchers continually find that age, gender, marital status, education level, family size, and social networks can explain the probability of migration (e.g., He and Gober 2003; Mu et al. 2021; Munshi 2020).[11] Spatial and social factors, such as geographic distance, urban amenities, public policies, and the Hukou registration barrier, also play a vital role in propelling or discouraging migration (e.g., Zhang et al. 2020). The academic debate on the weight of economic opportunities vs. amenities in migration has accordingly received more attention (e.g., Wang et al. 2020). Additionally, there is an increasing emphasis in the migration literature on environmental degradation, such as climate change and air pollution, caused mainly by industrialization and over-population (e.g., Liu and Yu 2020).

In addition to these determinants, the impacts of migration on migrants' families, such as consumption patterns and left-behind children (e.g., Li and Luo 2021; Meng and Xue 2020), impacts on areas of destination, such as skill aggregation, wage premiums,

[11]Although economists primarily focus on quantification, there exists a substantial body of migration literature in other disciplines that utilizes qualitative or mixed-method approaches to study social networks (e.g., Beck 2018; Liu et al. 2012).

and occupational upgrading (e.g., Chung et al. 2020; Zhao 2020), impacts on other potential migrants, such as networks and co-location (e.g., Foltz et al. 2020; Fu and Gabriel 2012), and impacts on the national economy if international movers are present, such as trade and foreign direct investment (Hatzigeorgiou and Lodefalk 2021), among others, have been widely discussed.

1.2.2 Expectations in Migration Decision-Making

Individual expectations about future outcomes are demonstrated as a compelling driving force behind migration decisions. For instance, De Jong (2000) found that income expectancies of remaining in home communities vs. living in alternative locations, in addition to residential satisfaction, were key determinants of migration intentions for both men and women in Thailand. Baldwin (2001) extended the core-periphery model, illustrating that the prevailing assumption of myopic migrants holds when migration costs are high, while forward-looking expectations arise in scenarios where migration costs are relatively low. More recently, Baumann et al. (2015) developed a Harris-Todaro model and used US state-level data to show that unemployment rates *per se* did not affect migration, but rather that changes in residents' expectations of unemployment across regions induced migration. Likewise, Shrestha (2020) conducted a randomized field experiment to demonstrate that individual expectations changed by gaining information on earnings and mortality rates abroad influenced actual migration decisions of potential migrants in Nepal, particularly of inexperienced migrants.

Other studies of note are of two types. The first is grounded in discrete choice models initiated by McFadden (1974), which provides a theoretical basis for gravity models of migration. Here, a canonical RUM model includes a deterministic and a stochastic component of utility and a time-specific migration cost. The distributional assumptions on the stochastic term determine the expected probability of selecting a destination. The deterministic term is typically modeled as a function of state variables, such as income, population, or temperature.[12] Bertoli et al. (2016) expanded the RUM model to allow migrants to make sequential decisions, such as return to their origin cities or move to alternative locations after migration. They assumed that migrants who are neither myopic nor living in a frictionless world where migration costs are zero respond to the future attractiveness and accessibility of all locations in the choice set. By using two proxies for expectations, they found that economic prospects at origin significantly influenced the scale of migration to Germany. Beine et al. (2019) also modified the RUM model but contended that agents' expectations of future outcomes were formed by the current level of economic activity and employment rates both within the country of origin and in a number of potential destination countries. Two proxies for signaling expectations about future employment probabilities at destination are found to have been influential in bilateral migration flows.

Another study type considers behavioral theories. Czaika (2015) developed the migration prospect theory grounded on Kahneman and Tversky's (1979) seminal work on prospect theory. Prospect theory suggests that the utility of an agent does not depend on an absolute outcome but on gains or losses gradually perceived relative to the reference point, or, in other words, on deviations from the status quo. By applying fundamental principles of prospect theory, i.e., loss aversion, reference dependence, risk preferences, and diminishing sensitivity, to the migration decision-making model, he demonstrated that migration flows responded more strongly to negative economic

[12] In other words, conventionally, the deterministic component is measured through instantaneous, absolute outcomes.

and unemployment prospects in home countries than to equal-sized positive prospects in Germany. Clark and Lisowski (2017) have also gained insights from prospect theory, emphasizing the endowment effect whereby people place a higher use value on the object they own than its market value to explain residential moving or staying. Their findings showed that, for internal migration in Australia, the probability of staying increases with stronger risk aversion, is higher for owners than for renters, and is higher still with longer duration at the current address.[13]

1.3 Theoretical and Empirical Framework

1.3.1 A Reference-Dependent Migration Model

When individuals are visionary, the location-specific utility that reflects the sequential nature of the decision-making problem defined by Bertoli et al. (2016) based on dynamic discrete choice models (e.g., Kennan and Walker 2011) is as follows:

$$U_{ijk,t} = w_{kt} + \beta A_{kt+1}(I) - c_{jk,t} + \epsilon_{ijk,t} \tag{1.1}$$

where $U_{ijk,t}$ is the utility of an individual i who moves from city j to city k at time t. w_{kt} is the deterministic instantaneous component of utility gained by moving from city j to city k at a time t. $c_{jk,t}$ describes the cost of moving from j to k at t. $\epsilon_{ijk,t}$ is an individual stochastic and time-specific serially-uncorrelated component of utility. Both w_{kt} and $c_{jk,t}$ are known to the individual. The expected utility gained by moving from city j to city k at time t and optimally choosing the preferred location from $t+1$ onward is $A_{kt+1}(I)$. We add I here to denote that the location preference hinges on the industry to

[13] Housing tenure, the duration at the current address, and neighborhood socioeconomic status are the variables used to construct endowment effects.

which job categories that individual i is searching belong. Jobs sought at destination are not necessarily identical to individual i's previous work at origin. $\beta \in [0,1)$ is the time discount factor of the expected utility. $\beta = 0$ represents the fact that potential migrants do not attach importance to future outcomes and as a result, make myopic decisions. It is also assumed that individual i chooses the preferred location after being aware of the stochastic component of utility at time t for all cities.

Beine et al. (2019) express the same idea that future outcomes matter, but assume that individual expectations are formed by extracting information from current economic conditions. Their model can be written as:

$$U_{ijk,t} = w_{kt} + \ln(E[A_{kt}(I)]) - c_{jk,t} + \epsilon_{ijk,t} \tag{1.2}$$

Although the model is constructed differently, $E[A_{kt}(I)]$ is basically the same to individual i as the discounted value $\beta A_{kt+1}(I)$ at time t, but will be different at some time $t+g \geq t+1$ if individual i moves away from k, because the expected instantaneous utility implies a permanent stay or move (Kennan and Walker 2011). Furthermore, Equation (1.2) takes the log of the expected utility to express the non-linearity and constant relative risk aversion.

Czaika (2015) also takes into account the non-linearity of utility and risk attitudes via replacing the utility function over absolute outcomes with the value function over gains and losses relative to a reference point, as does prospect theory. The reference-dependent migration value function derived from Kőszegi and Rabin (2009) can be written as:

$$V_{it} = V_{it}^k - V_{it}^j = M(\tilde{y}_{it}^k - \tilde{y}_{it}^j) + N(y_{it+1}^k, y_{it+1}^j | y_{it}^k, y_{it}^j) \tag{1.3}$$

$$where \;\; N(\cdot) = (y_{it+1}^k - y_{it}^k - y_{it+1}^j + y_{it}^j)^\alpha = (\Delta_{it+1}y^k - \Delta_{it+1}y^j)^\alpha \tag{1.4}$$

where V_{it} is the value of migration for individual i who moves from city j to city k at time t. $M(\cdot)$ is the regular component drawn from absolute outcomes in the origin city $\tilde{y}^j_{it}$ and the destination city $\tilde{y}^k_{it}$. In contrast, $N(\cdot)$ is the reference-dependent utility where present economic situations in the origin and destination city, i.e., y^k_{it} and y^j_{it}, respectively act as a reference point in adjusting present expectations about the future, i.e., y^k_{it+1} and y^j_{it+1}. The superscript α emphasizes the non-linearity.[14]

Here, we fine-tune the expected continuation payoff $A_{kt+1}(I)$ in Equation (1.1) to be reference-dependent, as follows:

$$A_{kt+1}(I) = y^k_{it+1} - y^k_{it} = \Delta_{it+1} y^k(I); A_{jt+1}(I) = y^j_{it+1} - y^j_{it} = \Delta_{it+1} y^j(I) \tag{1.5}$$

In a one-time migration scenario, we can rewrite Equation (1.5) as:

$$A_{kt+g}(I) = \frac{1}{\beta^g} E[A_{kt}(I)] = \frac{1}{\beta^g} E[A_{kt}(I)|r^g] \cdot E[r^g] = \frac{r^g}{\beta^g} \cdot \Delta_{it} y^k(I) \tag{1.6}$$

where $r \in (0,1)$ denotes the uncertainty between present trends and future realizations. $t + g \geq t + 1$ is any future point in time from time $t + 1$ onward. $r^g \Delta_{it} y^k(I)$ indicates that the further the future, the greater the uncertainty. Similarly, $\beta^g A_{kt+g}(I)$ can be understood as the further the future, the less influential will it be to the present utility.

Nevertheless, the continuation payoff $\Delta_{it+1} y^k(I)$ derived from Bertoli et al. (2016) entails a more intriguing intuition than the one-time approach that individuals can update their reference points after moving to k at time t and choose any alternative location q among the choice set D at time $t + 1$ in terms of their new reference points. If we assume that the stochastic component of utility follows an independent and identically distributed (i.i.d) Extreme Value Type-1 distribution (McFadden 1974) with

[14] In this chapter, we are only interested in the reference dependence, so for the description of other features, such as $N(\cdot)$ is concave (risk-averse) for expected gains when $N''(\cdot) \leq 0$ for $x > 0$ and convex (risk-friendly) for expected losses when $N''(\cdot) > 0$ for $x < 0$, please see Czaika (2015).

zero mean where τ is the Euler constant, the recursive form of the expected utility conditional on residing in k at time $t+1$, in terms of Kennan and Walker (2011) and Small and Rosen (1981), can be expressed as:[15]

$$\Delta_{it+1}y^k(I) = \tau + \ln\left(\sum_{q\in D} e^{w_{qt+1}-c_{kq,t+1}+\beta\Delta_{it+2}y^q(I)}\right) \tag{1.7}$$

Then, Equation (1.1) can be correspondingly rewritten as:

$$U_{ijk,t} = w_{kt} + \beta\left(\tau + \ln\left(\sum_{q\in D} e^{w_{qt+1}-c_{kq,t+1}+\beta\Delta_{it+2}y^q(I)}\right)\right) - c_{jk,t} + \epsilon_{ijk,t} \tag{1.8}$$

As McFadden (1974) shows, the probability of migrating from city j to city k can be estimated as:

$$P_{ijk,t} = Pr\{U_{ijk,t} = \max_{q\in D} U_{ijq,t}\} = \frac{e^{U_{ijk,t}}}{\sum\limits_{q\in D} e^{U_{ijq,t}}}$$

$$In\left(\frac{P_{ijk,t}}{P_{ijj,t}}\right) = w_{kt} - w_{jt} - c_{jk} + \beta\cdot[\Delta_{it+1}y^k(I) - \Delta_{it+1}y^j(I)] \tag{1.9}$$

where w_{jt} is the utility for individual i choosing to remain in city j at time t. The probability of k being chosen over j among the choice set D is equivalent to the probability of a binary choice of k over j if we assume the denominator is positive for all possible alternative choices.[16]

So far, we do not account for the non-linearity of utility as seen in Equations (2) and (4). It is very plausible that values attached to the migration project are not marginally linear but subject to the scale of change in economic situations, particularly, of the reference point. We can extend Equation (1.5) by taking the logarithm of $\Delta_{it+1}y^j(I)$ and

[15] For the reference-independent version, please see Bertoli et al. (2016).

[16] This assumption causes very little loss of generality. Please see McFadden (1974) for a detailed discussion.

$\Delta_{it+1}y^k(I)$. Thus, Equation (1.9) can be rewritten as:

$$In\left(\frac{P_{ijk,t}}{P_{ijj,t}}\right) = w_{kt} - w_{jt} - c_{jk} + \beta \cdot \ln \frac{\Delta_{it+1}y^k(I)}{\Delta_{it+1}y^j(I)} \tag{1.10}$$

1.3.2 Econometric Modelling and Techniques

Our innovative predictor is the main proxy for job prospects in the city of origin and city of potential destination, respectively. We use the term "trending" to highlight the fact that it captures upward and downward trends at the industry level. Macro surroundings often silently yet profoundly influence individual perceptions, and accordingly, trending signals derived here mirror the role of contextual evolution in forming expectations of all relevant individuals.

$$E\_Trending_{ij,t} = \frac{E_{ij,t} - E_{ij,t-1}}{E_{ij,t-1}} - \frac{E_{ij,t-1} - E_{ij,t-2}}{E_{ij,t-2}} = GR_{ij,t} - GR_{ij,t-1} = \Delta_t GR_{ij} \tag{1.11}$$

$$E\_Trending_{ik,t} = \frac{E_{ik,t} - E_{ik,t-1}}{E_{ik,t-1}} - \frac{E_{ik,t-1} - E_{ik,t-2}}{E_{ik,t-2}} = GR_{ik,t} - GR_{ik,t-1} = \Delta_t GR_{ik} \tag{1.12}$$

where the quantity of employment at time t in the sector of job categories that individual i looks for is $E_{ij,t}$ for the origin city j and $E_{ik,t}$ for the destination city k. In short, the trending indicator is the annual change in growth rates of industrial employment, i.e., $\Delta_t GR_{ij}$ and $\Delta_t GR_{ik}$. The beauty of this design is that positive growth does not necessarily lead to better employment prospects and vice versa. If we consider that $GR_{ij,t} = 1.5\%$ and $GR_{ij,t-1} = 1.8\%$, despite both being positive, the outcome is -0.3% signaling a slowdown. Likewise, for negative growth over two years, such as $GR_{ij,t} = -0.5\%$ and $GR_{ij,t-1} = -0.7\%$, $\Delta_t GR_{ij} = 0.2\%$ is still a positive value, because the scale of the present decline is narrower, implying that employment prospects stand a chance of getting better. In other words, individuals gain utility from staying or moving, not because of the growth *per se*, but due to whether progress is faster (better) or slower (worse) relative to the previous year at origin vs. destination.

The implication for this construction inspired by the migration prospect theory (Czaika 2015) can be linked to Baumann et al.'s (2015) findings that unemployment affects migration only if it alters expectations, and the central concept of Clark and Lisowski (2017). In our context, it could be understood that $GR_{ij,t} = 1.5\%$ does not trigger migration since no expected gain or loss emerges if $GR_{ij,t-1}$ also $= 1.5\%$. Moreover, Clark and Lisowski (2017) elaborate what prospect theory could offer for understanding why the majority of people prefer staying, because, in terms of loss aversion, people do not necessarily choose the highest expected utility gained via migration if they are more concerned about losing what they have. This offers insights into understanding our indicator from another angle. If we consider an industry with consistent growth at both origin and destination, to compete against the status quo bias ingrained in resident workers, deviations from two reference points are critical to add weight to moving and, consequently, to impel decision-making.[17]

Thus, we can formulate the distance of expected utility between the destination and origin city:

$$Distance(E\_Trending_{ijk,t}) = \Delta_t GR_{ik} - \Delta_t GR_{ij} \tag{1.13}$$

The distance variable corresponding to Equation (1.10) can be calculated as:

$$Distance^T(E\_Trending_{ijk,t}) = \ln(\Delta_t GR_{ik}) - \ln(\Delta_t GR_{ij}) = \ln(\frac{\Delta_t GR_{ik}}{\Delta_t GR_{ij}}) \tag{1.14}$$

where $Distance^T$ is the transformed version of Equation (1.13). The assumption is that

[17]We can elaborate "deviations from reference points" with two questions. First, are future situations at origin better than the present? Negative changes triggering loss aversion could strongly encourage emigration. Second, are changes at destination better than at origin? A larger scale of expansion at destination is required particularly when positive changes are also observed in origin cities.

holding the difference in employment prospects constant, potential migrants are more reluctant to move when their expectations of employment prospects in their origin cities are good, and the better the prospects, the less responsive potential migrants become (e.g., given $\Delta_t GR_{ik} - \Delta_t GR_{ij} = 1$, $\ln(2) - \ln(1) = 0.69$ vs. $\ln(10) - \ln(9) = 0.11$). Conversely, the worse the local prospects, the more susceptible potential migrants become.[18]

Likewise, we create a distance variable of reference-dependent wages as an additional proxy for job prospects:

$$Distance^T(Wage_{ijk,t}) = \ln(\Delta_t Wage_{ik}) - \ln(\Delta_t Wage_{ij}) = \ln(\frac{\Delta_t Wage_{ik}}{\Delta_t Wage_{ij}}) \tag{1.15}$$

where $\Delta_t Wage_{ik}$ and $\Delta_t Wage_{ij}$ are annual differences in average wages of staff and workers in city k and city j, respectively.[19] It is noteworthy that both our trending and wage indicators are the empirical counterpart of $\Delta_{it+1} y^j(I)$ and $\Delta_{it+1} y^k(I)$.

As binary outcomes cannot be log-transformed, our fixed-effects specifications act as a linear probability model approximation to Equation (1.10), which can be accurate in magnitude and sign for small parameter values in practical matters (McFadden 1974). The baseline to be estimated is as follows:

$$M_{ijk,t} = \alpha + \beta_1 X_{it} + \beta_2 Distance^T(Prospect_{ijk,t}) + \beta_3 Distance(Z_{ijk,t-1}) + \gamma_t + \epsilon_{ijk,t} \tag{1.16}$$

where the binary dependent variable (DV) $M_{ijk,t}$ equals 1 if individual i chooses to move from city j to city k at time t, and 0 otherwise. $Prospect_{ijk,t}$ could be either

[18] Because the trending indicator contains negative values, we applied the so-called "started logarithm" (Tukey 1977), i.e., $\ln(y + c)$ where $c > 0$ is set such that $y + c > 0$ for all y, to transform the variable. We estimated both versions and report the results of the untransformed distance predictor in the Appendix.

[19] It should be noted that this indicator measures general wage levels of a city, which does not vary by industry, because not all cities report statistics on sector-based wages and the measurement of such data among the rest is not consistent across cities throughout the sample period.

$E\_Trending_{ijk,t}$ or $Wage_{ijk,t}$. X_{it} is a row vector of individual and household characteristics. $Distance(Z_{ijk,t-1}) = Z_{ik,t-1} - Z_{ij,t-1}$ are city-level control variables. Because these covariates are state variables measured at the end of each year, we lag them by one period. γ_t are time-fixed effects and $\epsilon_{ijk,t}$ is an idiosyncratic error term.

In this chapter, the city covariates refer to income level, usually measured through GDP per capita (Beine et al. 2016), population density in relation to traffic congestion and informal settlement and slums (Tan et al. 2016), and urban amenities associated with two major aspects: the provision of healthcare (the number of hospital beds per person) and higher education resources (the number of higher educational institutions per 10,000 persons) (Czaika and Parsons 2017). To avoid multicollinearity, we replace GDP per capita with unemployment rates for estimating wage indicators. We also include the share of the tertiary sector and the number of enterprises above the designated size per 10,000 persons to account for gaps in regional commercialization and business density, respectively.[20] Additionally, in some specifications, we add the China Hukou Registration Index (CHRI) developed by Zhang et al. (2019) for 120 Chinese cities to control for Hukou entry policies and migration costs.[21] The individual and household covariates are: gender, marital status, Hukou type (rural or urban), self-evaluated health status, household income, age, years of schooling, and family migration network. Research suggests that the presence of networks lowers migration costs, increases the probability of migration, and is more important for the mobility of low-educated migrants than for the higher-educated (e.g., Beine et al. 2019). We construct it as a

[20] The term "enterprises above the designated size" was first used in China in 1996 and defined by the National Bureau of Statistics. Before 2011, firms with an annual output of 5 million Yuan or more were counted as enterprises above the designated size. The benchmark has raised up to 20 million Yuan since 2011.

[21] The CHRI divided the local Hukou registration policies into four overarching aspects – talent recruitment, general employment, investment and taxation as well as home purchase and then constructed the index for two stages (2000–2013 and 2014–2017).

dummy, indicating if there are any pioneer migrants within families prior to the move of each individual. Similar to Dai et al.'s (2019) consideration that the probability of an individual being connected to others increases with higher population density and such a higher probability can raise the number of network links and the probability of cross-connections among any three individuals, our variable captures both direct family networks and indirect migration networks connected through any migrant relatives.

We then add three fixed effects to Equation (1.16) to alleviate endogeneity issues, as follows:

$$\begin{aligned} M_{ijk,t} = \alpha + \beta_1 X_{it} + \beta_2 Distance^T(Prospect_{ijk,t}) + \beta_3 Distance(Z_{ijk,t-1}) \\ + \gamma_t + \gamma_j + \gamma_k + \gamma_s + \epsilon_{ijk,t} \end{aligned} \tag{1.17}$$

where γ_j, γ_k and γ_s are origin, destination and industrial sector fixed effects. Adding them helps reduce unobservable time-invariant or very slowly varying push and pull factors of each origin, destination, and industrial sector, such as the Hukou registration policies in origin or destination cities, industry-specific policies, or demographic characteristics of the population in each city or industry. This is the traditional strategy to control for multilateral resistance to migration in cross-sectional studies (Mayda 2010). In such cases, multilateral resistance to migration occurs when a destination city has different levels of attractiveness to people from the same place of origin due to gender, age, educational level, etc. This heterogeneity implies that origin-specific patterns of correlation across all potential locations exist in the stochastic component of utility.

However, the absence of future attractiveness of alternative locations which has an impact on choices of moving between j and k can still bias the estimation (Bertoli and Fernández-Huertas Moraga 2013). More specifically, assume that migration between Suzhou and Shanghai increases because the movers have worse expectations about their job prospects in Suzhou, and these worsening prospects are correlated with their worsening job prospects in Nanjing. Then, if the influence of alternative locations is not considered, the increase in the bilateral migration to Shanghai would be wholly attributed to worsening job prospects in Suzhou, resulting in overestimation. In other words, failure to account for multilateral resistance to migration might entail a violation of the IIA hypothesis underlying the discrete choice model discussed above.

Ideally, the solution to multilateral resistance to migration is Pesaran's (2006) common correlated effects (CCE) estimator if the cross-sectional and longitudinal dimensions of the panel are large enough. Yet our dataset does not meet this computing demand, so we follow the less data-demanding approach used in a wide range of applied migration literature (e.g., Beine and Parsons 2015; Ortega and Peri 2013; Royuela and Ordóñez 2018).

We change Equation (1.17), first by replacing monadic fixed effects of origin and destination with dyadic fixed effects of origin-destination, as follows:

$$\begin{aligned} M_{ijk,t} = \alpha + \beta_1 X_{it} + \beta_2 Distance^T(Prospect_{ijk,t}) + \beta_3 Distance(Z_{ijk,t-1}) \\ + \gamma_t + \gamma_{jk} + \gamma_s + \epsilon_{ijk,t} \end{aligned} \tag{1.18}$$

This specification is similar to the above but allows us to control for deterministic effects for each pair of cities. γ_{jk} can capture specific bilateral migration relationships between j and k, such as geographic distance, migration costs, or historic migration networks. Other variables remain the same, as above.

One of the most common approaches to dealing with multilateral resistance to migration is to apply dyadic fixed effects of origin-time, as follows:

$$M_{ijk,t} = \alpha + \beta_1 X_{it} + \beta_2 Distance^T(Prospect_{ijk,t}) + \beta_3 Distance(Z_{ijk,t-1}) \\ + \gamma_{jt} + \gamma_k + \gamma_s + \epsilon_{ijk,t} \tag{1.19}$$

where γ_{jt} is a vector of origin dummies for each year. All other variables remain the same as in the main specification. In terms of Ortega and Peri (2013) and Royuela and Ordóñez (2018), this method enables us to control for all the push determinants of migration decisions and particularly, multilateral resistance derived from heterogeneity in migration preferences that are constant across destination cities and that vary only by year and city of origin. Previous literature where DV was measured through migration flows also used γ_{jt} to account for the denominator $P_{ijj,t}$ (Beine et al. 2016).

Another common approach is to use dyadic fixed effects of destination-time in the place of origin-time, as follows:

$$M_{ijk,t} = \alpha + \beta_1 X_{it} + \beta_2 Distance^T(Prospect_{ijk,t}) + \beta_3 Distance(Z_{ijk,t-1}) \\ + \gamma_{kt} + \gamma_j + \gamma_s + \epsilon_{ijk,t} \tag{1.20}$$

where γ_{kt} is a vector of destination dummies for each year. All other variables remain the same as in the main specification. As explained in Beine and Parsons (2015) and Royuela and Ordóñez (2018), this strategy allows us to control for all the pull determinants of migration decisions and dynamic resistance derived from heterogeneity in the future attractiveness that are constant across origin cities and that vary only by year and city of destination.

Based on Equations (19)–(20), we replace the monadic fixed effects of industrial sectors with dyadic fixed effects of industry-time because our trending indicator rests on sectors:

$$\begin{aligned}
M_{ijk,t} &= \alpha + \beta_1 X_{it} + \beta_2 Distance^T(E_Trending_{ijk,t}) + \beta_3 Distance(Z_{ijk,t-1}) \\
&\quad + \gamma_{jt} + \gamma_k + \gamma_{st} + \epsilon_{ijk,t} \\
M_{ijk,t} &= \alpha + \beta_1 X_{it} + \beta_2 Distance^T(E_Trending_{ijk,t}) + \beta_3 Distance(Z_{ijk,t-1}) \\
&\quad + \gamma_{kt} + \gamma_j + \gamma_{st} + \epsilon_{ijk,t}
\end{aligned} \tag{1.21}$$

where γ_{st} is a vector of industry dummies for each year, capturing heterogeneity in migration preferences that vary by year and industrial sector, such as emerging jobs.[22] All other variables remain the same as in the main specification.

As we have seen, the less data-demanding solution to multilateral resistance to migration relies on utilizing various structures of fixed effects. Nevertheless, a fixed effects estimator can be significantly biased in non-linear models. Beck (2018) demonstrates that the larger the number of fixed effects, the stronger bias imposed on fixed effects logit models. In contrast, the critique of using a linear probability model (LPM) is usually twofold: a) predicted probabilities might be negative or above 1, such as 1.2 or -0.4, that are unrealistic, and b) the dichotomous DV renders heteroskedasticity which violates one of the OLS assumptions that all disturbances have the same variance. Yet the heteroskedasticity can be addressed with heteroskedasticity-consistent robust standard errors, while the first concern is most often why the LPM is not preferred. Horrace and Oaxaca (2006) shows that the bias and inconsistency of LPM increase with a greater proportion of predicted probabilities falling outside the unit interval. In other words, if all predicted probabilities fall between 0 and 1, the linear probability estimator can be unbiased and consistent.

In our sample, 8.3% (6.2%) of observations present a negative predicted migratory

[22] Since the wage indicator is not sector-specific, we do not estimate this group of equations for it.

probability drawn from Equation (1.17). We, thus, report estimates of the sub-sample where all predicted probabilities are within the unity as a supplement. Further, we consider a multilevel logistic regression as part of the robustness check. It enables us to perform logistic regression along with controlling for unmeasured context-specific influences on potential migrants that is in accord with the theoretical perspective that people from the same place of origin tend to behave more similarly than do individuals from other places due to a variety of spatial and socio-cultural proximity.

Let $\pi_{it} = Pr(M_{ijk,t} = 1)$, our two-level logistic random intercept model is formulated as follows and purely an empirical counterpart of Equation (1.10):

$$\log(\frac{\pi_{it}}{1-\pi_{it}}) = \alpha + \beta_1 X_{it} + \beta_2 Distance^T(Prospect_{ijk,t}) + \beta_3 Distance(Z_{ijk,t-1}) + \gamma_t + \mu_C \tag{1.22}$$

where μ_C is assumed to be i.i.d normally distributed with a zero mean and level-2 variance σ_C, accounting for the effects of being in city group C on the log-odds that $M_{ijk,t} = 1$. β_1 is still level-1 unknown parameters, and β_2, β_3 and β_4 are level-2 parameters to be estimated. The index C can be either an origin city or a destination city. Later, we estimate both.

We can extend the two-level model to a three-level model if we postulate that within the same city group, the type of contextual information that people access and the way they are impacted differ across education levels. In this case, city effects μ_C become level-3 and a new level-2 random intercept, i.e., μ_{Cv}, is added to Equation (1.22). The index v denotes the year of schooling.

In addition, Wintoki et al. (2012) show that the generalized method of moments (GMM) estimation yields better results than a fixed effects estimator for at least two potential issues of endogeneity: unobserved heterogeneity and simultaneity. On the basis of Hansen (1982), the two-step system GMM is further developed by Blundell and Bond (1998) where lagged first differences are utilized as instruments in the level equation at the cost of an additional assumption that first differences of instrument variables are uncorrelated with unobserved unit-specific heterogeneity. We, thus, further continue our analysis using a system GMM estimator.[23] The model is almost the same as Equation (1.16) but with an inclusion of industry fixed effects.

1.3.3 Data Description

Data Source

The quasi-panel is created by combining a nationally representative cross-sectional micro-data, the 2017 China Household Finance Survey (CHFS), with city-level longitudinal statistics during 1997–2017, retrieved from the China Data Institute. This approach is not uncommon for applied research.[24] The CHFS is one of the very few nationwide surveys that allow us to identify where migrants are from and where they settle at the city level and the year in which their migration occurs.[25] After comprehensive data cleaning and compiling, our panel eventually contains 10,254 migrants and 56,173 natives. For details of the data preparation, please see Appendix D.

It is worth noting again that all individual and household variables were retrieved from the 2017 wave. After merging the cross-sectional survey data with the longitudinal statistics, we re-calculated the age and years of schooling of each individual for each year and only retained observations of ages between 16 and 65. As a result, we finally

[23] It is unreasonable to assume that the previous migration decision affects the present migration decision. We, therefore, do not consider adopting a lagged DV.

[24] To give an example, Schmidt-Catran and Spies (2016) exploited hybrid methods to investigate cross-sectional as well as longitudinal effects of migration on native Germans' support for welfare.

[25] Another option is the CMDS, however, we do not have permission to access it.

have three time-varying variables at this level, that is: age, years of schooling, and family migration network. As migrants are defined as those who have moved across prefecture boundaries, and to focus on the decision-making problem, our binary DV is only equal to 1 in the year of migration and to 0 at all other points in time. Furthermore, we distinguished individuals who had moved and transferred their Hukou to destination cities from natives and specified them as migrants. As migrants in China are often measured by the separation of the Hukou and resident places, people who transferred the Hukou to their cities of destination are treated as natives in the majority of migration studies (Zhang et al. 2020). However, ignoring those migrants with Hukou transfer leads to an underestimate of the scale of migration.

Descriptive Statistics

In our RUM model, we discuss that the expected utility $A_{kt+1}(I)$ is conditional on the industry in which job categories that individual i seek are mainly based. From the CHFS, we learn each individual's industry group, if he or she has a formal job. Table 1.1 classifies the industry groups based on the Chinese national standard number "GB/T 4754" and presents the corresponding statistics of each group. As shown, at the aggregate level, the primary sector has the lowest proportion of labor, as opposed to the tertiary sector. Among sub-sectors, 21.3% of individuals in our sample were employed in Manufacturing, followed by Construction (9.7%), Social Services (9.1%), and Wholesale and Retail Trade (9%).

The trending indicator is eventually quantified by the statistics of the three main sectors. Approximately 25,000 natives and 3,000 migrants in our sample do not provide industry groups of their jobs because they are non-employee workers. Thus, total employment statistics are used for this group instead of sectoral statistics. Accordingly, we have four sets of industry-specific dummies applied in our estimation (primary, secondary, tertiary, and total). Furthermore, Figure 1.4 is the geographic distribution of emigrants employed in the secondary sector moving across province boundaries, from which we see five Chinese provinces suffered the greatest labor outflux: Sichuan, Hubei, Anhui, Hunan, and Guangdong. Similarly, Figure 1.5 is the plot looking at the tertiary sector. Here, Guangdong still sent out the most migrants followed by Sichuan, Hubei, Anhui, Hunan, and Liaoning. In Figure 1C.1, we also graphically present where migrants are from, defined by the prefecture boundary.

Table 1.1: Industrial Classification & Statistics

Industry Group	Num. of Obs.	Main Sector
Farming, Forestry, Animal Husbandry	2695	Primary Sector
Mining and Quarrying	488	Secondary Sector
Manufacturing	7765	Secondary Sector
Electric Power Gas and Water Production and Supply	1184	Secondary Sector
Construction	3517	Secondary Sector
Wholesale and Retail Trade	3294	Tertiary Sector
Transportation Storage Post and Telecommunications	2531	Tertiary Sector
Hotel and Catering Services	1959	Tertiary Sector
Information Transmission, Software and Information Technology	870	Tertiary Sector
Banking and Insurance	1015	Tertiary Sector
Real Estate	466	Tertiary Sector
Leasing and Business Services	333	Tertiary Sector
Scientific Research, Technical Service and Geologic Prospecting	234	Tertiary Sector
Management of Water Conservancy, Environment and Public Facilities	455	Tertiary Sector
Social Services	3314	Tertiary Sector
Education	2112	Tertiary Sector
Health, Social Security and Social Welfare	1562	Tertiary Sector
Culture, Sports and Entertainment	613	Tertiary Sector
Public Management and Social Organisation	1942	Tertiary Sector
Total	36349	

Source: Authors' calculation using CHFS.

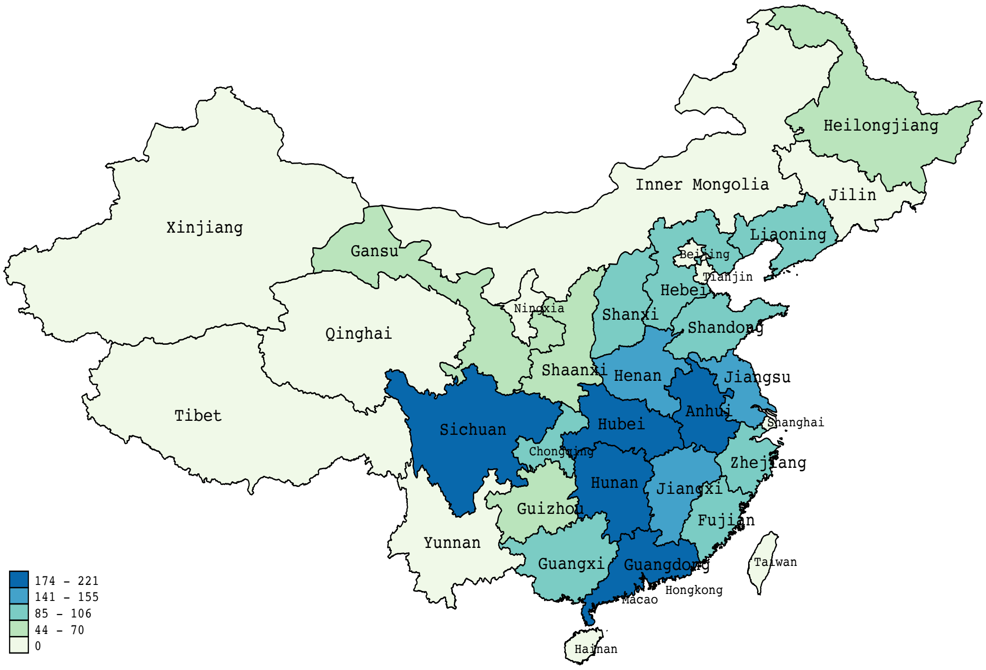


Figure 1.4: Province boundary map – emigration distribution (secondary sector).

Source: Author's elaboration using CHFS.

Figure 1.5: Province boundary map – emigration distribution (tertiary sector).

Source: Author's elaboration using CHFS.

Table 1.2 contains descriptive statistics of the survey data and city statistics, respectively, as well as the DV and the trending indicator ($Distance^T(E\_Trending_{ijk,t})$)

in the final quasi-panel.[26] More specifically, migrants in our sample were on average 11 years younger, had 1.4 more years of schooling, and obtained 6,800 Yuan more in annual wages after-tax than did natives. The gender ratio of females to males among natives was 46:54, but 42:58 among migrants. As to the Hukou type, 44% of natives, as opposed to 34% of migrants, had an urban Hukou.[27] Family sizes did not seem to differ between natives and migrants (1.87 versus 1.89), but fewer migrants were in a marital relationship. Self-evaluated health status shows that migrants overall evaluated their health as slightly better than that of natives.

Moreover, compared to natives, migrants were mostly from economically underdeveloped cities where the GDP per capita was on average 4,600 Yuan less in 2000 and 12,500 Yuan less in 2010 than for natives' resident cities. Generally, migrants moved to economically developed cities where the GDP per capita was 9,800 Yuan more in 2000 and 50,000 Yuan more in 2010 than their areas of origin. Likewise, destination cities usually had 9% higher ratios of the tertiary sector to the total than cities of origin and three to four times more enterprises above the designated size per 10,000 persons. These destination cities were also on average 41% more densely populated than migrants' cities of origin but with better public medical and higher education resources per capita. In addition, natives' resident cities applied more stringent Hukou registration policies: the CHRI was 0.17 higher before 2013 and 0.28 higher since 2014 than in migrants' cities

[26] We acknowledge several limitations here. Firstly, as a quasi-panel, most of the individual and household variables are time-invariant. Although the CHFS has four waves, about one-third of migrant samples were new to 2017, and earlier waves still do not have information prior to 2011. Secondly, the migration history we can learn is incomplete. It is possible for migrants to move several times within a year, however, we can only identify their most recent destination. Thirdly, industries to which individuals' jobs belong depend on their most recent employment. A mismatch between what we learn from the 2017 wave and actual examples that individuals evaluated is possible. Otherwise, examining job prospects at the sub-industry level would be even more interesting.

[27] Prefecture cities in China usually consist of several districts and counties. They have dual functions of administering both rural and urban areas. An increasing population with the rural Hukou lives in urban areas because of local urbanization and within-prefecture migration (Song 2014).

of origin. However, the Hukou barrier did not indeed prevent migrants from moving to higher indexed cities.[28] This is mainly because cities setting higher registration barriers were, in general, more economically developed (Zhang et al. 2019). Lastly and importantly, annual differences in wage levels in destination cities were, in general, 3.6% higher than in origin cities (3353.42 versus 3237.40 Yuan) and increased to 27.6% (4241.59 versus 3322.94 Yuan) at the time that migration occurred. Similarly, although the percentage change in growth rates of industrial employment in origin cities fared slightly better than destination cities, for example, 4.7% at origin vs. 3% at potential destination in 2000 and 1.6% at origin vs. 0.9% at potential destination in 2010, the situation completely reversed at the time that migration occurred: 0.8% at origin vs. 2.2% at destination, on average.

[28] Actually, their cities of destination were on average 0.37 higher before 2013 and 0.49 higher since 2014 than their cities of origin.

Table 1.2: Descriptive Statistics

Variable	Description	Obs.	Mean	Std. Dev.	Min	Max
CHFS 2017						
migrant	If native=0; migrant=1	75,099	0.17	0.38	0	1
gender	Female=0; Male=1	75,099	0.55	0.50	0	1
age	Age in 2017	75,099	47.82	17.66	16	85
education	Educational level: no schooling=1, PhD=9	74,894	3.48	1.75	1	9
marriage	Married=1; Otherwise=0	72,835	0.77	0.42	0	1
hukou_type	Urban Hukou=1; Rural=0	74,961	0.41	0.49	0	1
health_status	Health degree: very good=1, very bad=5	75,072	2.52	1.03	1	5
hh_income	Log of household income	75,099	13.89	0.36	-9.21	15.61
family_size	The number of family members living together	12,214	1.88	1.28	1	10
wage	After-tax wage last year	24,736	37533.76	45066.41	20	3100000
City Statistics						
all_GR	Growth rates in all sectors	5,522	0.0208	0.4284	-0.9673	18.36
prima_GR	Growth rates in the primary sector	5,230	0.1581	3.55	-1	120.96
second_GR	Growth rates in the secondary sector	5,738	0.0132	0.4543	-0.9267	22.72
tertiary_GR	Growth rates in the tertiary sector	5,738	0.0117	0.3121	-0.9395	7.78
ppDen	Log of population density	5,667	5.85	1.01	1.55	9.55
lngdppc	Log of GDP per capita	5,492	9.73	0.9582	7.01	12.58
coop	The number of enterprises above the designated size per 10,000 persons	5,425	2.53	3.32	0.12	36.29
medical	Number of beds in hospitals per person	5,705	0.0033	0.0016	0	0.0189
highEdu	Number of higher educational institutions per 10,000 persons	5,739	0.0155	0.0194	0	0.1197
tertiaryRatio	The ratio of output values of the tertiary sector	5,780	0.6529	0.1553	0.2029	0.9897
CHRI	The Hukou registration stringency index	2,167	0.6614	0.3370	0.1331	2.63
Merged						
migrate	=1 at the time of migration; =0 before and afterwards	1,128,624	0.0084	0.0914	0	1
distance_ETrend	The ratio of employment prospects at destination to origin	969,998	-7.37e-06	0.0045	-0.6976	0.4884
distance_wage	The ratio of wage prospects at destination to origin	1,050,567	0.0387	0.2940	-5.94	7.75
pioneer	If any pioneer migrants within families =1; No=0	1,128,624	0.0237	0.1522	0	1

Notes: Approx. 10,000 observations in the 2017 CHFS survey cannot match city statistics because their resident cities are not included in the statistics. 1,466 households answered with zero or even negative annual income due to debts and losses on investment, among others. The variable “distance_ETrend” and “distance_wage” are the transformed, as defined in Equations (14) and (15). The statistical currency is Yuan.
Source: Authors’ elaboration using CHFS, CHRI, and China Data Institute (2021).

1.4 Estimation

1.4.1 Results

Results in terms of Equations (16)–(21) are reported in Table 1.3. The most basic result in Column (1) indicates that a 10% change in the ratio of employment prospects (annual changes in sector-specific employment growth) at destination to origin is positively associated with an increase of 0.876 percentage points in migratory probabilities. It is worth noting that small R^2 values here are endemic to and a direct result of our discrete choice setting. Then, we added individual, household and city covariates to the second model, as seen in Column (2). The effects of employment prospects increase considerably, from 0.876 to 1.933 percentage points, and additionally, having any pioneer migrants within families prior to the move leads to an increase of 6.67 percentage points in migratory probabilities. In addition, we estimate the same models using the distance in employment growth rates, i.e., $GR_{ik,t} - GR_{ij,t}$, and find that it has little impact on migration, as opposed to the trending indicator.[29]

From Column (3) onward, we begin to eliminate unobserved heterogeneity where various options of fixed effects were added, and in Column (4), we included the CHRI (henceforth, the Hukou index). As reported in Column (3) and Columns (5)–(9), the magnitude of the coefficients of the ratio of sector-based employment prospects in cities of destination to cities of origin is between 1.321 and 1.788 percentage points. Similarly, the magnitude of the coefficients of the family migration network is between 5.96 and 6.66 percentage points. Column (4) can be seen as a subpopulation analysis because only 120 cities have Hukou indices. Results indicate that, after controlling for the Hukou registration stringency, the influence employment prospects exert on migration

[29] We show the major results in Table 1B.1.

climbs to 1.636 percentage points. And a unit increase in the distance in Hukou indices raises the probability of migration by 2.22 percentage points. The stringency gap *per se* would not attract labor but projects regional disparities in economic development and urbanisation.[30] However, the interaction effect between the Hukou index and employment prospects is just statistically significant at the 10% level.

[30] The top four indexed cities are Beijing, Shanghai, Guangzhou, and Shenzhen, colloquially known as "Bei-Shang-Guang-Shen", representing the most developed areas in China.

Table 1.3: Determinants of Migration Decisions (1997–2017): Employment Prospects with Fixed Effects

No Restrictions

	OLS		Multilateral Resistance to Migration						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
distance_ETrend	0.0876***	0.1933***	0.1408***	0.1636***	0.1491***	0.1721***	0.1321***	0.1354***	0.1788***
	(0.0297)	(0.0290)	(0.0330)	((0.0579)	(0.0329)	(0.0576)	(0.0321)	(0.0314)	(0.0576)
distance_CHRI				0.0222**					
				(0.0087)					
distance_(JobTrendXCHRI)				0.2212*					
				(0.1129)					
pioneer		0.0667***	0.0619***	0.0576***	0.0666***	0.0596***	0.0614***	0.0617***	0.0601***
		(0.0043)	(0.0035)	(0.0063)	(0.0045)	(0.0033)	(0.0036)	(0.0037)	(0.0034)
Ind. Controls		Y	Y	Y	Y	Y	Y	Y	Y
City Controls		Y	Y	Y	Y	Y	Y	Y	Y
Constant	0.0090***	0.0622***	0.0603***	0.0406***	0.0396***	0.0607***	0.0600***	0.0601***	0.0606***
	(0.0009)	(0.0045)	(0.0041)	(0.0038)	(0.0042)	(0.0042)	(0.0042)	(0.0042)	(0.0041)
Time FE	Y	Y	Y	Y	Y				
Industry FE			Y	Y	Y	Y	Y		
Origin FE			Y	Y		Y			Y
Destination FE			Y	Y			Y	Y	
Pairs of cities FE					Y				
Origin-year FE							Y	Y	
Dest-year FE						Y			Y
Industry-year FE								Y	Y
R^2	0.0014	0.0459	0.0534	0.0509	0.0842	0.0721	0.0697	0.0696	0.0721
Obs	969998	749219	729965	408877	729960	729769	729725	748981	749034

Notes: Complete results of columns (1)–(9) are reported in Appendix A using the untransformed trending indicator. Standard errors shown in parentheses are clustered at the destination city. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Source: Created by authors using CHFS, CHRI, and China Data Institute (2021).

Table 1.4: Determinants of Migration Decisions (1997–2017): Supplement

Predicted Probabilities between 0 and 1							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
distance_ETrend	0.1515***	0.2562**	0.1438***	0.1947**	0.1281***	0.1312***	0.2185***
	(0.0454)	(0.1291)	(0.0480)	(0.0771)	(0.0431)	(0.0428)	(0.0776)
distance_CHRI		0.0292***					
		(0.0097)					
distance_(JobTrendXCHRI)		0.1933					
		(0.1952)					
pioneer	0.0625***	0.0564***	0.0666***	0.0610***	0.0626***	0.0628***	0.0618***
	(0.0035)	(0.0059)	(0.0046)	(0.0033)	(0.0036)	(0.0036)	(0.0033)
Ind. Controls	Y	Y	Y	Y	Y	Y	Y
City Controls	Y	Y	Y	Y	Y	Y	Y
Constant	0.0621***	0.0431***	0.0386***	0.0596***	0.0614***	0.0613***	0.0589***
	(0.0042)	(0.0042)	(0.0041)	(0.0043)	(0.0042)	(0.0042)	(0.0042)
Time FE	Y	Y	Y				
Industry FE	Y	Y	Y	Y	Y		
Origin FE	Y	Y		Y			Y
Destination FE	Y	Y			Y	Y	
Pairs of cities FE			Y				
Origin-year FE					Y	Y	
Dest-year FE				Y			Y
industry-year FE						Y	Y
R^2	0.0564	0.0558	0.0878	0.0801	0.0750	0.0748	0.0804
Obs	668644	306227	724519	680253	678702	696327	701926

Notes: Here, regressions are conditional on the predicted probabilities produced by the models of Columns (3)–(9) reported in Table 1.3. Standard errors shown in parentheses are clustered at the destination city. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Source: Created by authors using CHFS, CHRI, and China Data Institute (2021).

In Table 1.4, we restricted observations to those whose predicted probabilities are within the unity. Except for Column (2), all regressions involved more than 90% of observations. Estimates of employment prospects in Columns (1), (2), (4) and (7) are found to be larger than their counterparts reported in Columns (3), (4), (6) and (9) of Table 1.3. In contrast, Columns (3), (5) and (6) show slightly smaller effects of employment prospects on migration decisions compared to Columns (5), (7) and (8) of Table 1.3. In sum, the magnitude of the coefficients of the ratio of sector-based employment prospects at destination to origin is between 1.281 and 2.185 percentage points. Furthermore, the interaction term is statistically insignificant. Given these results, we conclude that our linear probability estimator performs acceptably.[31]

[31] After computing the predicted probabilities of each model, we counted the proportion of migrant observations with negative fitted values. The proportions are 1.1%, 0.9%, 0.7%, 2.4%, 1.4%, 1.4%, and 2.4% of the total.

Regarding estimates of control variables reported in Table 1A.1, we see that migratory probabilities increase for people who are male, unmarried, younger, or more educated. As we focus on labor migrants, our findings are distinguished from those studies that do not primarily analyze labor migration. For example, in this study, men are found to be more likely to migrate than women, whereas the opposite story is not uncommon among other migration scenarios, such as permanent migration (e.g., Meng 2020; Zhang et al. 2020). As for city covariates, the distance in income is found statistically insignificant in Columns (3)–(9) when we further control for heterogeneity in migration preferences. Instead, two variables are consistently significant, that is, the distance in the provision of healthcare and business density. As expected, abundant medical resources positively drive migration, while the negative role of business density in migration seems surprising.[32]

In Table 1B.2 we report the results of wage prospects. Columns (1)–(4) show that a 10% change in the ratio of wage prospects at destination to origin is positively correlated with an increase of 0.237, 0.041, 0.039, and 0.033 percentage points in migratory probabilities. However, once we include both origin and destination fixed effects, the wage indicator turns statistically insignificant. This manifests why including bilateral fixed effects are important here because, by using only origin (destination) city fixed effects, certain time-invariant or slowly varying unobserved factors across cities on the other side could still be correlated with the wage indicator. Considering wage prospects are not consistently statistically significant, we do not perform robustness checks for it in the next section.

[32] With a simple regression, the sign remains negative.

1.4.2 Robustness Checks

Multilevel Logistic Regression

Results of multilevel logit models are reported in Columns (1)–(6) of Table 1.5. In Column (1), we consider cities of origin as the higher level to control for origin-specific factors affecting the probability of migration. As we see, the intra-class correlation coefficient (ICC) is merely 0.0432, indicating that observations within the same place of origin are different from each other, so we switched to Columns (2) and (3) where cities of destination and origin-destination pairs are treated as level-2, respectively.[33] Here, the ICC is 0.3355 for destination and 0.2783 for pairwise cities, presenting substantial evidence of clustering, and their coefficients are smaller than in Column (1). In other words, observations within the same destination (origin-destination pairs) have a much lower degree of variability compared to their origin-nested counterparts.[34] As destination-nested models present the largest ICC, we mainly interpret these results.[35] Because coefficients in the logistic regression are not as easily interpreted as coefficients in the linear regression, we graphically illustrate marginal effects of the second model with an interval of 0.2 in Figure 1.6a, holding all other variables at mean. The plot clearly shows an increasing trend that, when the ratio of employment prospects at destination to origin gets larger, its effects on motivating migration become increasingly stronger.

It has been widely demonstrated that education levels have a considerable impact on the propensity for migration and location preferences (e.g., Fu and Gabriel 2012; Meng 2020). Thus, we treated years of schooling as a new level, subordinate to cities. In other

[33] ICC is calculated as the ratio of the between-group variance relative to the total variance in the sample. It describes the extent to which observations within city groups are similar to each other.

[34] This is the reason, in addition to the CHFS's sampling design, for us to cluster standard errors at the destination city for all non-hierarchical models (see Colin Cameron and Miller 2015).

[35] The pairwise nested model produces results most similar to the fixed-effects models but has a lower ICC than the destination-nested model.

words, people with identical educational attainments nested in the same locations are supposedly much more similar to each other than their fellow migrants who received a higher or lower level of education. As can be seen, the ICC becomes slightly larger in Columns (4)–(5) than in Columns (1)–(2), whereas almost no change arises from adding the level of education under pairwise nests. It should be noted that cities are level-2 in Columns (1)–(3), but level-3 in Columns (3)–(6). Hence, the variance between cities is 0.1484, 1.6611, and 1.2684 in two-level logit models and 0.1054, 1.6046, and 1.2685 in three-level logit models. Interestingly, the distinction between education groups is even greater than the extent to which that the grouping of origin can account for, as opposed to Column (6), where the variance between level-2 groups within the same origin-destination pairs is nearly zero. The fifth model gives an in-between result: its level-2 variance is around one-seventh of the level-3. We further plot its marginal effects, holding all covariates at mean. As seen in Figure 1.6b, the marginal effects are also continuously increasing.

In Table 1A.2, we report the complete regression results of Table 1.5. The results of covariates vary depending on the definition of nests, except for the distance in income, which is consistently statistically significant across all models. The distance in the provision of healthcare (business density) is found positive (negative) at the 5% (10%) level in pairwise nested models but not in others. In contrast, the distance in population density and the share of the tertiary sector exhibit positive impacts on increasing the probability of migration.[36]

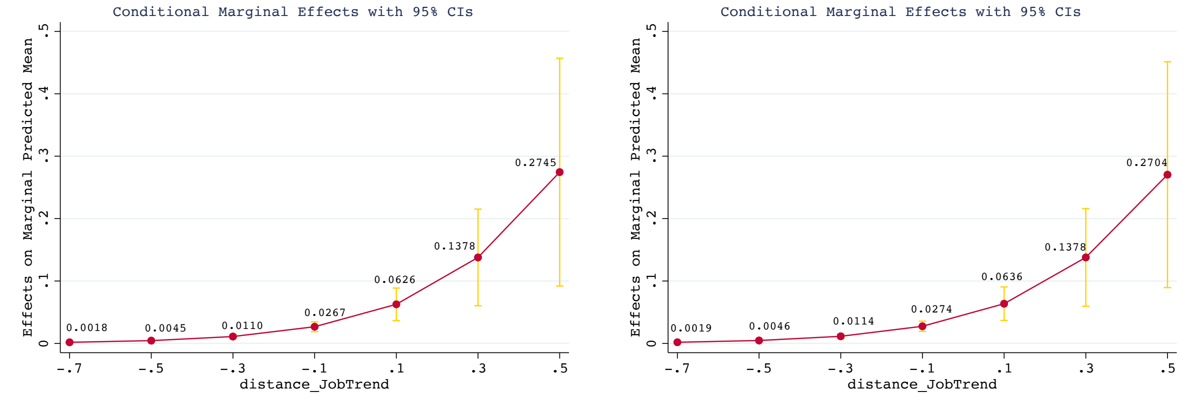
[36]The relationship between the distance in the share of the tertiary sector and migration decisions is negative when we account for origin- and destination-specific effects simultaneously as in pairwise nested models, despite being statistically insignificant.

Table 1.5: Determinants of Migration Decisions: Multilevel Logit and Two-step System GMM

	Two Level Logit			Three Level Logit			GMM		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
distance_ETrend	5.3903***	4.5964***	3.3363***	5.3922***	4.5847***	3.3363***	0.3157***	0.2493***	0.2751***
	(1.0279)	(0.6027)	(1.0040)	(1.0181)	(0.6120)	(1.0040)	(0.0754)	(0.0569)	(0.0582)
Ind. Control	Y	Y	Y	Y	Y	Y	Y	Y	Y
City Control	Y	Y	Y	Y	Y	Y	Y	Y	Y
Intercept	-1.839***	-1.5841***	0.0077	-1.9504***	-1.7681***	0.0077	0.0610***	0.0590***	0.0590***
	(0.3101)	(0.2573)	(0.2468)	(0.3137)	(0.2857)	(0.2468)	(0.0046)	(0.0046)	(0.0046)
Level 2 var.	0.1484	1.6611	1.2684	0.1869	0.2416	1.99e-33			
	(0.0372)	(0.1671)	(0.1009)	(0.0312)	(0.0455)	(3.38e-34)			
Level 3 var.				0.1054	1.6046	1.2685			
				(0.0398)	(0.1685)	(0.1009)			
ICC	0.0432	0.3355	0.2783	0.0816	0.3595	0.2783			
	(0.0104)	(0.0224)	(0.0160)	(0.0112)	(0.0214)	(0.0160)			
Nest	origin	destination	pair	origin	destination	pair			
sub-Nest				education	education	education			
Time FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Industry FE							Y	Y	Y
Num. of instruments							182	180	201
AR(2)							0.724	0.790	0.790
Hansen's J test							0.349	0.263	0.404
Difference-in-Hansen tests									
GMM instruments for levels – Excluding group							0.114	0.108	0.154
GMM instruments for levels – Difference (null H = exogenous)							0.451	0.355	0.578
GMM instrument for distance_trend – Excluding group							0.520	0.354	0.526
GMM instrument for distance_trend – Difference (null H = exogenous)							0.114	0.192	0.173
Obs	749219	749219	749219	749219	749219	749219	729965	729965	729965

Notes: In Appendix A, we report complete results using the untransformed trending indicator. AR(2) is the Arellano-Bond test for second-order serial correlation with the null hypothesis of no serial correlation in disturbances. Hansen's J test is a test of over-identifying restrictions (in other words, the overall validity of the instruments) using the J statistic of Hansen (1982). Likewise, the difference-in-Hansen test is designed to test the validity of subsets of the instruments. Endogenous and predetermined variables are instrumented with their corresponding second- and third-order lags in columns (5)–(6). Three more orders of lags are added in column (7). The trending indicator, the income (GDP per capita), and the share of the tertiary sector are treated as endogenous in column (5), while other city-level covariates are treated as predetermined (not strictly exogenous). In columns (6)–(7), the population density and business density are additionally treated as endogenous. Moreover, individual and household covariates are specified as exogenous. The Windmeijer correction (Windmeijer 2005) is used in the GMM estimation, and corresponding standard errors are clustered at the destination city. Besides, robust standard errors are also applied in multilevel logit models.* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Source: Created by authors using CHFS and China Data Institute (2021).



(a) Two-level random intercept model. (b) Three-level random intercept model.

Figure 1.6: Marginal effects of the destination-nested model.

Notes: All results are statistically significant at the 1% level.
Source: Created by authors using CHFS and China Data Institute (2021).

Two-step System GMM

In Columns (7)–(9) of Table 1.5, we report two-step system GMM estimates with different lag and instrument strategies. We adopted two types of fixed effects: time and industry. The former is aimed at absorbing any instant shock exposed to all units, while the latter is concerned to ensure that variations of employment prospects are estimated within sectors. The first GMM model is the baseline where we utilized two orders of lags to instrument the endogenous and predetermined variables. We initially treated the trending indicator, the income, and the share of the tertiary sector as strictly endogenous, while all other city covariates were treated as predetermined. The coefficient of the trending indicator shows that a 10% change in the ratio of sector-based employment prospects at destination to origin causes an increase of 3.157 percentage points in migratory probabilities. Then, we applied the same lag strategy but, additionally, treated the population density and business density as endogenous variables. As a result, the coefficient is still statistically significant at the 1% level but relatively smaller than the previous result. We kept this instrument strategy but added three more orders of lags to instrument our variables. Adding higher orders of lags can help remove serial correlation but may also impose the weak instrument problem, so we did not include more lags. We learn from Column (9) that the magnitude of the coefficient of our trending indicator is a bit larger but falls exactly between the values of coefficients in Columns (7) and (8). Moreover, the results here are larger than the fixed effects estimates reported in Table 1.3 and, interestingly, the model of Column (9), where we included the Hukou index and confined observations to having a within-unity predicted probability, produces the closest estimate (2.562 percentage points in Table 1.4).

As seen in Table 1A.2, half of the city-level control variables are still statistically insignificant in Column (7). While in Columns (8)–(9), the majority appears to be statistically significant, as opposed to the business density, which was initially found to have an impact, but then became negligible when it was treated as endogenous. The second and third GMM models are better than the first in dealing with unobserved heterogeneity, their results are thus more reliable. According to these results, we find that the distance in income, the provision of healthcare, the provision of higher education and population density have positive effects on driving migration. In contrast, the distance in the share of the tertiary sector negatively affects migratory probabilities. The relationship between the distance in population density and migration uncovered here can be attributed to the fact that migrants are attracted to large cities that inevitably are densely populated (Chen and Fan 2016). For the latter, a possible explanation is that considering, in our sample, 39% of total migrants worked in the tertiary sector, cities with a relatively lower share of the tertiary sector are more likely to present a faster pace of growth.[37]

[37] Based on the variance inflation factor, the colinearity between the trending indicator and the share of the tertiary sector is very weak (their VIFs are 1.01 and 3.01, respectively). Also, the Pearson's correlation coefficient is just 0.0021.

The results of Hansen's J test for over-identification are well above 0.25, a threshold suggested by Roodman (2009), but far from 1, pointing to a 34.9%, 26.3% and 40.4% chance of a type one error if the null is rejected, and no issue of instrument proliferation. The null hypothesis of the Arellano–Bond test is not rejected, indicating no second-order serial correlation in disturbances.[38] By further checking the difference-in-Hansen test for the validity of subsets of instruments, the differenced models are evidenced as dynamically complete, implying the instruments used in the level models are valid. We also report test results of GMM instruments of the trending indicator, from which we can conclude that its corresponding specified instruments used in the level models are exogenous.

1.5 Conclusion

In this chapter, we investigated the effects of job prospects on individual migration decisions across Chinese prefecture boundaries. To this end, we assembled a unique quasi-panel based on the 2017 China Household Finance Survey and the prefecture city statistics between 1997 and 2017. By accounting for city-level bilateral variations in parallel with individual and household characteristics, we filled gaps in the existing Chinese migration literature in two aspects: a) migration decisions at the city level are quite scarce due to data limitations, and among them, individual and household characteristics are always absent in analyzing regional longitudinal effects, and b) the previous models that have controlled for these characteristics rely on micro-data whereby regional factors are either missing, monadic, or at the province level. Equally importantly, Chinese migration research is desperately lacking in visionary migration scenarios. By

[38] Despite not being reported, no serial correlation was found in AR(3)–(4).

constructing proxy variables for wage and employment prospects, respectively, we enriched economic incentives of labor migration from a forward-looking angle. Further, we theoretically added to the literature of migration decision-making by synthesizing the virtues of dynamic discrete choice modeling framework and reference dependence derived from prospect theory. Following it, we drew corresponding empirical specifications and applied various monadic and dyadic fixed effects to address multilateral resistance to migration. Further, we considered multilevel logistic regression and two-step system GMM estimation for the robustness check. In sum, this study, acclimatized to the new migration pattern that provincial-level migration has been less popular, deepened the understanding of relationships between regional employment structures and labor mobility in both level and scope.

Wage prospects are influential to migration decisions when we control for unilateral fixed effects, while results become statistically insignificant once bilateral fixed effects are involved. Thus, our primary findings are that a 10% increase in the ratio of employment prospects in cities of destination to cities of origin raises the probability of migration by 1.281–2.185 percentage points, and the effects tend to be stronger when the scale of the ratio is larger. Having a family migration network causes an increase of approximately 6 percentage points in migratory probabilities. Additionally, labor migrants are more likely to be male, unmarried, younger, or more educated.

Our results align with the existing global literature on the influence of individual expectations of future outcomes, bilateral socio-economic distance, and deviations from reference points on migration (e.g., Baumann et al. 2015; Bertoli et al. 2016; Czaika 2015). Further, they contribute to the ongoing academic debate surrounding the conflicting results related to economic opportunities and urban amenities (e.g., Czaika and Parsons

2017; Wang et al. 2020). While some recent studies emphasize urban amenities as the key driver, our findings, partly consistent with those of Wang et al. (2020),[39] suggest that economic attractiveness remains a more crucial factor for Chinese labor migrants. We also add to the discourse on migration networks (e.g., Meng and Xue 2020; Munshi 2020). As limited by data, this study could not incorporate social networks measured at the community level which may weaken the impact of family networks on migration if households are tied to established large community networks (Winters et al. 2001). As a whole, our findings suggest that small- and medium-sized cities can benefit from concentrating resources to accelerate employment growth in certain sectors and thus build "comparative advantages" in talent attraction and retention. On one hand, by adopting this strategy, cities can outperform others in attracting some types of migrants and laying the foundation for generating spillover effects. On the other hand, amidst China's slowdown (e.g., Chen and Groenewold 2019), lower growth rates coupled with a range of factors like exorbitant housing prices in large cities discourage immigration, thus opening opportunities for small- and medium-sized cities.

[39] Our findings show that medical resources, rather than the provision of higher education, are influential to migration decisions, while Wang et al.'s (2020) results are opposite.

Chapter 2

Intra-city mobility dynamics and commuting behaviors surrounding the Zero-COVID policy and reopening in China

2.1 Introduction

As the first country to experience the outbreak of COVID-19, China strictly implemented the Zero-COVID Policy during January 2020–December 2023. During that period, a series of containment measures, including travel restrictions, mass testing, and the QR code health e-passport, were introduced, exerting an unprecedented influence on travel behaviors. Since the State Council announced the "10-point plan" on December 07, 2022, China has shifted towards reopening rapidly. For instance, the 2023

Chunyun saw 1.595 billion passenger trips, 50.5% more than the same period in 2022.[1]

Human mobility is the crux of both the Zero-COVID Policy and the current reopening. Kraemer et al. (2020) find that mobility data can precisely explain the initial spread of COVID-19 in China. On the one hand, curbing mobility effectively mitigated the growth of COVID-19 cases (e.g., Fang et al. 2020; Glaeser et al. 2022). On the other hand, unfastening travel restrictions can mobilize economic activities (Spelta and Pagnottoni 2021). The trade-off between reduced and enhanced mobility involves balancing health outcomes and economic growth. Wu et al. (2023) find that China's lockdown can explain 2.8 percentage points of its GDP loss, with decreased labor mobility being an important channel. At the same time, intra-regional mobility was found to be the main mechanism through which regions with more economic activities experience higher infection rates of COVID-19 in India (Chakraborty and Mukherjee 2023). As cities return to pre-pandemic life, regional disparities in various aspects related to human mobility, including health and economic performances, may widen due to differing levels of exposure to and resilience against COVID-19 across cities.

While COVID-19 literature on human mobility has mostly focused on its associations with demographic or socio-economic characteristics (e.g., Hu et al. 2021; Long and Ren 2022), travel restrictions (e.g., Fang et al. 2020; Gibbs et al. 2020), COVID-19 spread (e.g., Iacus et al. 2020; Tokey 2021), carbon emissions (e.g., Lei et al. 2022; Liu et al. 2020b), and, more recently, work-from-home (e.g., Cicala 2022; Delventhal et al. 2022). Mobility dynamics surrounding how daily local travel (within cities) shifts towards normalcy, which can reflect various aspects of recovery of individual lives around the end of the Zero-COVID Policy, is however not yet examined. Similarly, previous studies on the

[1]Chunyun, also called "Spring Festival Travel Rush", is a period with extremely high travel rates in China around the Chinese New Year. During such periods, family reunions are often the primary driving force behind trips (Li and Ma 2022).

Zero-COVID Policy have investigated its consequences including health, wellbeing, interpersonal trust, human mobility, and job-housing relationships (e.g., Chen et al. 2023; Fang et al. 2023; Mu et al. 2023), research that considers the risk-level system, a key component of the Zero-COVID Policy, and its impact on local travel behaviors is non-existent, due primarily to data limitations. In fact, prior to this chapter, only Gong et al. (2023) showcases the impacts of being classified as risk areas on a series of economic indicators including population inflows and outflows.

Therefore, the first objective of this chapter is to assess the transition paths of local travel in Chinese cities. Specifically, the aim is to identify if regional gaps in human mobility are widening, and if so, which cities are falling behind. To this end, we compile a novel panel using intra-city mobility data derived from Baidu Maps,[2] a Chinese equivalent to Google Maps,[3] and COVID-19 risk-level data in terms of the State Council of the PRC's release (Gong et al. 2023). Further, the second objective is to disentangle the impacts of the "high-risk" alert on different travel behaviors. Liu et al. (2021) point out that intra-city mobility is negatively correlated with socio-economic development levels. Lessons drawn from unraveling the differentials can enable us to have a more in-depth look at regional differences in life dynamics during the pandemic.

Overall, in a combination of spatiotemporal and difference-in-difference (DiD) analyses, we a) analyze human mobility trajectories of 368 Chinese cities surrounding the strict periods of the Zero-COVID policy and the end of it and b) examine the effects of exposure to high COVID-19 risk in the city on essential (work-purpose) and non-essential (dining-, leisure- and recreational purpose) travel rates.[4] To this end, we use

[2]We base our analysis on intra-city mobility because it is more directly related to economic activities in the city (e.g., Chakraborty and Mukherjee 2023; Liu et al. 2021), while inter-city travel can often be converted into intra-city travel as long as the inbound passengers do not just pass by.

[3]In contrast to Google mobility data, one virtue of the Baidu mobility data is that it allows changes to be compared between provinces and cities.

[4]The State Council of the PRC classified the COVID-19 risk into three levels. Areas that never have

three unique indicators: intra-city travel strength, home-workplace (HW) commuting rates, and dining, leisure, and recreational (DLR) travel rates. In sum, our approaches provide an overview of similarities and differences in mobility patterns, suggesting that instead of one-size-fits-all policies, post-pandemic travel policymaking can be developed and promoted among the identified clusters. In addition, we show that lock-downs or analogous measures had much more profound impacts on essential travel than what we can literally learn from the documents, signifying the importance of developing more follow-up policies to make up for the toll when such interventions become necessary again.

Our key findings are fourfold. Firstly, we look at the moving paths of the intra-city travel strength during January 17, 2022–March 12, 2023, by performing Phillips and Sul's (2007) clustering algorithms.[5] In terms of the mobility dynamics during the studied period, six clusters are detected, where the last cluster, representing the lowest level of intra-city travel strength, is falling significantly behind the others. It overall suggests that gaps, at least mobility-wise, are widening between clusters, but not between all the cities. In addition, we uncover a division between Western China and the rest of the country, with the majority of cities in Qinghai, Tibet, and Xinjiang dropping well below the average.

We then assess the moving paths of the HW and DLR travel during May 17, 2021–June 26, 2022 using the same clustering method.[6] Our second finding is that essential

confirmed cases or do not have new confirmed cases in the last 14 consecutive days are deemed as low-risk areas. Medium-risk areas are those that have no more than 50 confirmed cases and no clustered outbreak in the last 14 consecutive days. Likewise, when there are more than 50 confirmed cases or any clustered outbreak, high risk will be assigned. We use the term "exposure to high COVID-19 risk" to describe if cities have any areas identified as high risk.

[5]This method enables us to trace the relative trajectory of each city to the cross-country average in a dynamic manner. We further discuss its advantages in the next section.

[6]It is worth noting that the sample period studied here is consistent with the later DiD analysis. For the consideration of causal analysis, we choose June 26, 2022 as the termination date since China released the Protocol for Prevention and Control of COVID-19 (Edition 9) on June 27, 2022 to relax its Zero-COVID

travel can be classified into nine clusters and three divergent cities of which moving patterns are incompatible with any clusters. The relative distance between six out of the nine clusters tends to increase. A divide is found as well, with 41 western cities exhibiting higher rates of essential travel than the rest of China. Yet the divide does not exist for non-essential travel, which comes to our third finding that the results of DLR travel are much more balanced across cities with a grouping of six and four divergent cities. In particular, the transition paths of all clusters and cities, except for Hong Kong, are quite consistent over time.

Lastly, we investigate the impacts of exposure to high COVID-19 risk on HW and DLR travel. Two things are especially noteworthy. First, as long as a city was alerted to having a high-risk area, commuting rates there dropped significantly, and the declines persisted. Second, the impacts on different travel behaviors were uneven, being greater to essential than to non-essential travel.

This chapter contributes to three strands of literature. First, it is the first study to analyze the changing regional differences in human mobility during the transitional period. Research on spatiotemporal mobility has still been extremely scarce in the literature, and among the very few studies related to COVID-19, regional difference, despite being somewhat reflected on maps, is never examined in a dynamic manner. Further, research on regional mobility predominantly emphasizes the early phase of the pandemic (e.g., An et al. 2023; Hu et al. 2022; Tokey 2021) with objectives to study or predict the co-evaluation of COVID-19 spread and on micro-mobility tied to particular

policy in various aspects, while the eighth edition was published on May 14, 2021. One of the changes in the ninth edition is to reduce the duration of containment measures implemented in high-risk areas from 14 days to 7 days. In our DiD setting, we consider having at least one area to be identified as high risk in the city as receiving the treatment. Thus, we choose May 17, 2021 and June 26, 2022 as the starting and ending dates to ensure that the implemented Zero-COVID policy is consistent throughout the studied period.

travel tools (e.g., Hu et al. 2021; Jiao et al. 2022; Xin et al. 2022), such as bike sharing and/or e-scooters. In addition, the closest study to this research is Mu et al. (2023) where inter- and intra-city Baidu Mobility Data are used to analyze new mobility patterns in 2021 and the effects of experiencing at least one local confirmed case on mobility levels relative to the benchmark.[7] While this study is distinguishable from it and other existing literature in a range of dimensions, including measure, unit, period, and approach.

Second, it adds to the broader body of research on COVID-induced socio-economic outcomes. Specifically, there are some papers looking into the effects of COVID-19 transmission or travel restrictions on mobility and/or economic activities (e.g., Caselli et al. 2021; Fang et al. 2020; Li et al. 2021). In addition, studies on job commuting and work-from-home have been emerging (e.g., Cicala 2022; Delventhal et al. 2022; Mitze and Kosfeld 2022). Complementing both, this is also the first study to exploit the "high risk" alarm set by the Zero-COVID policy to investigate changes in commuting behaviors.

Further, there is an increasing number of works assessing different aspects of convergence for the Chinese economy using the Phillips and Sul (2007, 2009) approach, such as Bai et al. (2021), Tian et al. (2016), Valerio Mendoza et al. (2022), and Zhu and Lin (2020). This is the first research to apply this sophisticated econometric technique, previously employed in the empirical convergence literature for longer horizons, to assess short-term variations in regional patterns on a weekly basis.

[7] They find a so-called "localized mobility pattern" that inter-city (intra-city) mobility was 16% (9%) lower (higher) than January 4–10, 2020 (the benchmark). Further, declines in both inter- and intra-city mobility are found to be larger for less developed cities, which is inconsistent with that of Liu et al. (2021). This inconsistency could be a result of the difference in their benchmark data.

2.2 Data description

2.2.1 Baidu Mobility Data

Intra-city mobility data are retrieved from Baidu Qianxi, a program launched in 2014 built on Baidu map's location-based service (LBS). As one of the most popular web mapping applications in China with 130 billion real-time records of location service requests per day, over 500 million people using its AI-powered voice assistant, and supporting more than 500,000 mobile apps to locate their users, it enables Baidu Qianxi to precisely trace and outline population movements within and between cities and provinces on a daily basis. Baidu mobility data have been increasingly used in recent research, such as Fang et al. (2020), Gibbs et al. (2020), Huang et al. (2022), Liu et al. (2021), and Mu et al. (2023).

To perform Phillips and Sul's (2007, 2009) log(t) test, we normalize the indexation outcomes of the intra-city travel strength and HW and DRL travel provided by Baidu Qianxi and analyze them on a weekly basis.[8] A summary is given as follows:

- Intra-city travel strength index: accounts for the average ratio of the number of people with trips in the city to the total population in the city from Monday to Sunday.
- HW commuting index: accounts for the average share of home-workplace trips in the city from Monday to Sunday.
- DLR travel index: accounts for the average share of dining, leisure, and recreational trips in the city from Monday to Sunday.

[8] As Baidu Mobility Data can take any numeric value greater than or equal to 0, without normalizing them, it is hard to compare results between sub-samples straightforwardly.

2.2.2 COVID-19 risk-level data

The COVID-19 risk-level data are collected from the website of the State Council of the PRC and compiled on a daily basis with 368 prefecture cities (Gong et al. 2023). For each city or equivalent administrative unit, there are three indicators recording whether it had low-, medium-, and high-risk areas, respectively, in terms of the COVID-19 risk-level system. Among the three levels, low-risk areas were the relatively "peaceful" places with basic preventive measures, such as wearing masks and body temperature testing in public areas. More measures including quarantine were conducted in medium- and high-risk areas, while the latter underwent more stringent containment, including the shutdown of all the places that are populated and with high population mobility, such as theaters, libraries, public transportation, gyms, shopping malls, supermarkets, etc., break of elective operations in hospitals, and stopping risky population (medical workers, cleaners, couriers, migrants, and service staff, among others) from working in high-risk areas.

We also aggregate the risk-level data to be weekly and use the high-risk indicator in our DiD analysis because high-risk areas experienced the most extensive lockdown. We base our DiD analysis primarily on HW travel but consider DLR travel a good complementarity to it because both of them relate to economic activities with DLR travel more specific to the service sector.[9] Thus, we define the high-risk indicator to be 1 if the city has at least one high-risk area during Monday–Friday, and 0 otherwise. In other words, if a high-risk area appears on Saturday or Sunday, the indicator will be coded 1

[9] As Baidu data are anonymized and aggregated, we cannot identify individual travel purposes. Individuals who work in restaurants should be recorded in essential travel, yet in non-essential travel in fact, if they do not mark the location of their homes and workplaces in the app, and vice versa. Further, the data could be biased towards younger people and longer-distance local trips and against underdeveloped cities.

only from the following week.[10] The reason behind is rather simple – people usually do not commute for work on weekends. As aforementioned, when this indicator equals 1, we interpret it as there is an exposure to high COVID-19 risk in the city. Eventually, 50 cities had been exposed to high COVID-19 risk during our studied period. In contrast, 316 cities that had never experienced high risk are treated as control groups. In Table 2.1, we present the descriptive statistics for all the variables.

Table 2.1: Descriptive Statistics

Time Frames	Description	Mean	Std. Dev.	Min	Max
Baidu, $T = 60$					
Jan 17, 2022–Mar 12, 2023	Intra-city travel strength	0.3794	0.0872	0	1
Baidu, $T = 58$					
May 17, 2021–June 26, 2022	HW commuting rates	0.3071	0.0830	0	0.8319
May 17, 2021–June 26, 2022	DLR travel rates	0.0979	0.0507	0	1
Risk level, $T = 58$					
May 17, 2021–June 26, 2022	high risk of COVID-19	0.19	0.3908	0	1

Notes: Authors' elaboration using the Baidu Mobility Data and COVID-19 risk-level data (Gong et al. 2023).

2.3 Methodologies

2.3.1 Clustering algorithms

We employ an empirical framework developed by Phillips and Sul (2007, 2009) to study human mobility dynamics at the city level in China. While the econometric technique proposed by Phillips and Sul (2007) was originally developed with the aim to examine convergence patterns over the long run, it is equally suitable for transition modeling for shorter periods, given its non-restrictive time series properties. In particu-

[10]Consistent with the staggered adoption design, also known as event study designs, assumed in the staggered DiD literature (including Callaway and Sant'Anna (2021)), we conceptualize our treatment setting as one in which cities are unlikely to completely forget the "memory" of being exposed to high COVID-19 risk, at least in the short term. This means that there is a "scarring" effect in affected cities even after the "high risk" alert has been lifted. Moreover, we excluded one city that was treated in the first period.

lar, the main advantage of this approach over existing tests that analyze co-movements in the variables studied is that it does not depend on any stationarity assumptions, and therefore it does not require the time series to be cointegrated. In fact, it is reminiscent of an asymptotic cointegration test that does not suffer from the small sample problems and other limiting characteristics of conventional unit root and cointegration testing. This in turn allows the model to admit a wide range of transition dynamics, and moreover, the different types of city-specific individual trajectories can be visually examined. In addition, as part of the methodology, an iterative clustering procedure enables us to endogenously identify subgroups within the panel, without assuming *a priori* club classification of cities.

Hence, we adopt the Phillips and Sul (2007) approach to analyze human mobility and cluster Chinese cities based on inter- and intra-city travel dynamics, i.e., move-in and commuting rates. Specifically, we define a nonlinear time-varying factor model as

$$X_{it} = \delta_{it}\mu_t, \tag{2.1}$$

where X_{it} represents the log of Baidu move-in rates or commuting rates for city i at period t. Our variable of interest can be decomposed into a common trend component μ_t and a time-varying idiosyncratic loading parameter δ_{it}. The latter includes information about the individual trajectory of city i relative to the common trend μ_t, and any departure of city i from this trend depends on how individual, city-specific mobility differences within the panel change over time. The loading coefficient δ_{it} thus provides a measure of the relative distance between X_{it} and μ_t.

To formulate our null hypothesis, we model the idiosyncratic factor loadings δ_{it} in a semi-parametric form, as suggested by Phillips and Sul (2007):

$$\delta_{it} = \delta_i + \frac{\sigma_i}{L(t)t^{\alpha}}\xi_{it}, \tag{2.2}$$

where δ_i is fixed, ξ_{it} is $i.i.d.(0,1)$ across i but weakly dependent over t, σ_i are idiosyncratic scale parameters, and α is the decay rate, i.e., the speed at which city-specific differences decrease over time. This representation enables us to test whether $\frac{\sigma_i}{L(t)t^{\alpha}}\xi_{it} \to 0$ in Equation 2.2 as $t \to \infty$ for any $\alpha \geq 0$, in which case $\delta_{it} \to \delta$, suggesting that differences in mobility characteristics across cities disappear over time. Specifically,

$$H_0: \qquad \delta_i = \delta \text{ for all } i \text{ and } \alpha \geq 0,$$

which is tested against the alternative:

$$H_A: \qquad \{\delta_i = \delta \text{ for all } i \text{ with } \alpha < 0\} \text{ or } \{\delta_i \neq \delta \text{ for some } i \text{ with } \alpha \geq 0, \text{ or } \alpha < 0\}.$$

The null hypothesis implies common behavior across all cities in the panel, whereas the alternative encompasses two possible outcomes: *(i)* the existence of city clubs, where one or more subgroups of cities get aligned in terms of mobility patterns over time, with possibly one or more diverging units, and *(ii)* divergence of all cities in the panel.

In order to test the null hypothesis, Phillips and Sul (2007) define the following parameter h_{it}:

$$h_{it} = \frac{X_{it}}{N^{-1}\sum_{i=1}^{N} X_{it}} = \frac{\delta_{it}}{N^{-1}\sum_{i=1}^{N} \delta_{it}}, \tag{2.3}$$

also referred to as the relative transition path, which measures the relative departure of the loading coefficient δ_{it} of city i at time t from the cross-sectional mean. In other words, the parameter h_{it} traces out the individual trajectory of each city i in relation to the panel average. The transition paths of the 369 Chinese cities in our panel may either approach each other or exhibit – transitory or persistent – deviating patterns over the sample period studied. The null hypothesis cannot be rejected when all cities move toward the common trend, i.e., $h_{it} \to 1$ for all i as $t \to \infty$, in which case the cross-sectional variance of h_{it} decays asymptotically:

$$H_t = \frac{1}{N}\sum_{i=1}^{N}(h_{it} - 1)^2 \to 0 \qquad \text{as} \qquad t \to \infty. \tag{2.4}$$

Equation 2.4 is formally tested using the $\log(t)$ test:

$$\log\left(\frac{H_1}{H_t}\right) - 2\log L(t) = a + b\log(t) + u_t, \tag{2.5}$$

for $t = [rT], [rT] + 1, \ldots, T$, where $r > 0$.[11] Specifically, we run a one-sided t-test for $\alpha \geq 0$ using the estimate $\hat{b} = 2\hat{\alpha}$ and heteroscedasticity and autocorrelation consistent (HAC) standard errors, and the null hypothesis is rejected at the 5% significance level if $t_{\hat{b}} < -1.65$. If the null is rejected for the overall sample, a clustering algorithm is applied based on repeated $\log(t)$ tests and a set of criteria, in order to detect all city clubs as well as units that deviate from the rest within the panel.[12]

2.3.2 DiD with multiple periods

The emergence of COVID-19 outbreak and accordingly high-risk areas is quite random, resulting in great variation in treatment timing across cities. Further, treatment effects can hardly be constant between cities and over time because of the varying levels of cities' resilience against COVID-19 and lockdowns and the loosening of containment measures once the "high-risk" alert has been downgraded. Thus, we exploited Callaway

[11] Phillips and Sul (2007) recommend a slowly varying function $L(t) = \log(t)$ and setting $r > 0$ on the interval $r \in [0.2, 0.3]$ for sample sizes $T \geq 100$ and $T \leq 50$, respectively.

[12] The reader is referred to Phillips and Sul (2007, 2009) for a detailed description of the clustering algorithm.

and Sant'Anna's (2021) DiD with multiple periods estimator to address treatment effect dynamics and heterogeneity in a unified manner.[13] The group-time average treatment effect is defined as follows:

$$ATT(g,t) = E[Travel_t(g) - Travel_t(0)|Risk_g = 1] \tag{2.6}$$

where $Risk_g$ is a binary variable equal to 1 if cities are first exposed to high COVID-19 risk in time period g, for g=2, ..., τ; $Travel_t(0)$ denote the potential essential or non-essential travel rates at time t if cities remain unexposed throughout the studied period; $Travel_t(g)$ capture the potential outcome if cities were to be exposed to high COVID-19 risk in period g for the first time. Following the aggregation schemes discussed in Callaway and Sant'Anna (2021), we then estimate more aggregate causal parameters by summarizing the ATTs across different groups g, at different points in time t, and across different lengths of exposure, $e = t - g$.

2.4 Empirical results

2.4.1 Intra-city travel dynamics

Figure 2.1 presents the clustering results of the intra-city travel strength during January 17, 2022–March 12, 2023. The dashed blue line in Figure 2.1(a) represents the cross-country average mobility. As shown, the moving paths of the 368 cities form six clusters with two of them performing above the average. Two out of the six clusters are above-average, containing 255 (69%) cities. Here, two things are of note. Firstly, all the

[13]It is worth noting that as we use those never-treated as controls, the estimator is identical to Sun and Abraham's (2021) cohort-and-period specific estimator. In addition, although De Chaisemartin and d'Haultfoeuille (2020) provides a more general setup, it focuses on the instantaneous treatment effect. Since a high-risk area can be as small as a building or neighborhood in the city, its impacts are likely for city-level mobility data to take a while to capture.

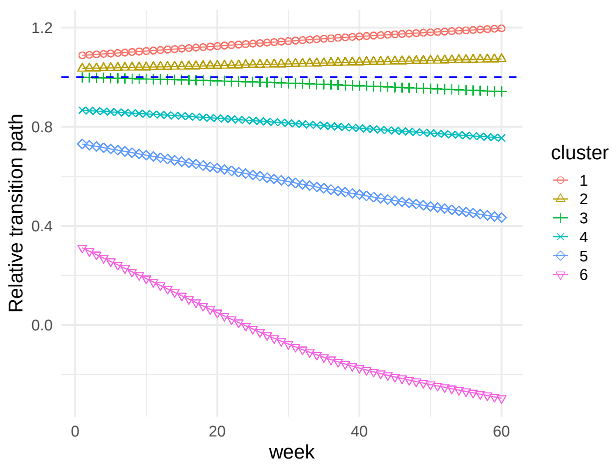
clusters are drifting apart, suggesting that the distance in intra-city mobility between clusters tends to increase over time. Secondly, Clusters 5 and 6, accounting for 30 cities (8.2%), are falling significantly behind the others. According to Table 2A.1, cities in these clusters include Lanzhou, Xining, Lhasa, and Ürümqi, the provincial capitals of Gansu, Qinghai, Tibet, and Xinjiang. Despite being less developed, Jinchang (in Gansu; its 2020 GDP is 35.9 billion Yuan) and Alar (in Xinjiang; its 2020 GDP is 33.2 billion Yuan) in Clusters 2 and 1 are found to have more intensive local travel than that of Lanzhou (its 2020 GDP is 288.7 billion Yuan) and Ürümqi (its 2020 GDP is 333.7 billion Yuan). As for Qinghai and Tibet, intra-city mobility in most of the cities there is a little higher than that of their provincial capitals (Xining in Cluster 5; Lhasa in Cluster 6).

Further, since cities in Cluster 6, which is the most worrisome group compared to the others, are mostly in Xinjiang and Tibet, these two provinces are likely to also lag behind in the recovery of economic activities.[14] Socio-economic development levels, including infrastructure, may explain part of the low mobility, but cannot account for the declining paths observed here. While a plausible explanation is the increasing population outflows in these regions, such as migrant workers who move to other provinces,[15] at the time that the Zero-COVID policy comes to an end.

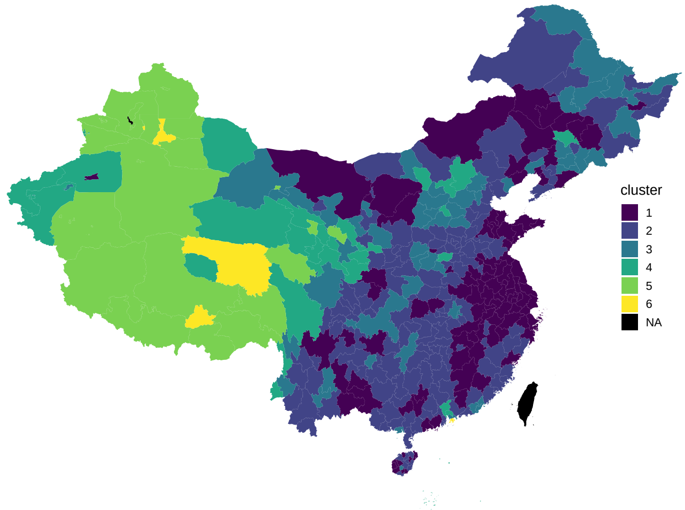
Moreover, the map in Figure 2.1(b) highlights a clear division between Western China and the rest of the country. In particular, the majority of cities in Qinghai, Tibet, and Xinjiang fall into Clusters 4–6. In contrast, cities outside the two provinces are primarily grouped in Clusters 1–3, with just a few exceptions such as Hohhot, Guangzhou, and Dongguan in Cluster 4. Although reasons behind lower mobility could substan-

[14]As aforementioned, population density does not influence the results as our indicators are comparable between cities.

[15]For instance, on Mar 12, 2023, 18.7% of passengers moved from Qinghai to Lanzhou, 12.8% from Gansu to Xi'an, 16.4% from Tibet to Chengdu, and 8.9% from Xinjiang to Jiuquan. However, these results do not mean that outbound passengers outnumber inbound passengers there.



(a) Transition paths.



(b) Geographical distribution.

Figure 2.1: Spatiotemporal results of intra-city travel strength during January 17, 2022–March 12, 2023.

Notes: The result of each city can be found in the Appendix and our dashboard.
Source: Created by authors using the Baidu Mobility Data.

tially vary across cities, we can draw some common behaviors from visualizing the average local travel strength for each cluster. As seen in Figure 2A.1, intra-city mobility dropped more substantially and persistently in Clusters 4–6, compared to Clusters 1–3, starting from around Aug 7th, 2022 until December 11, 2022. A plausible explanation is the spread of COVID-19 in China during those days. We also plot the daily new confirmed COVID-19 cases in China from Jul 1st, 2022 to Dec 23rd, 2022 (the last date of update) in Figure 2A.2. A surge in cases appeared right after August 1, 2022, and continued to climb until a few days before December 11, 2022. Further, as the dashed cyan lines show, lower clusters tend to take longer to return to the previous levels. On the other hand, Clusters 1–3 experienced sharper declines than Clusters 4–6 during the second half of December, they however recovered within only 2–3 weeks. We also find that Macao and Hong Kong are in Clusters 5 and 6, but the fewer intra-city trips recorded by Baidu might be because fewer Baidu Maps users there.

2.4.2 HW & DLR travel dynamics

Figure 2.2 presents the results of HW (essential) and DLR (non-essential) travel during May 17, 2021–June 26, 2022. The transition paths of HW commuting are sorted into nine clusters with three cities diverging from all the others: Changchun, Macao, and Xi'an. Here, HW commuting in Macau is the highest, pulling away from all the other clusters. In contrast, the other two cities are diverging below Cluster 8. As Figure 2.2(a) shows, the movement of all the clusters and divergent cities, except for Cluster 6, tends to fan out more or less. Among them, cities in Clusters 7–9 experience commuting increasingly less than the national average. The division between Western China and the rest of the country is also evident in Figure 2.2(b), although it is not as distinctive as in Figure 2.1(b) because the results are more scattered here. Most of the western cities are in Clusters 1–4 (red), in contrast to Clusters 6–9 (blue) which are the most common among other regions. While both parts have some areas falling into Cluster 5 (grey).

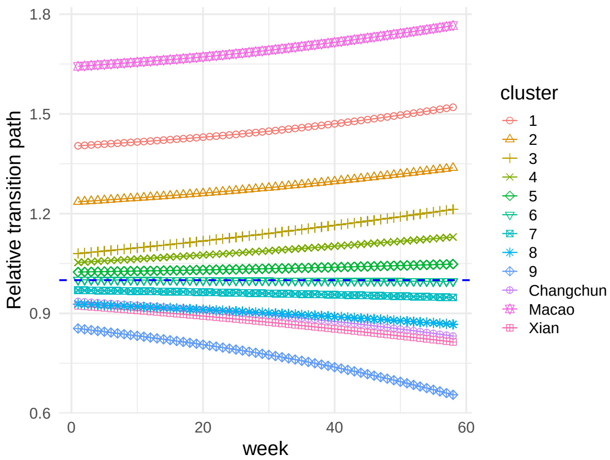
The division as a whole could be because no city in Clusters 1–3 was alerted to have high-risk areas during the studied period, as seen in Figure 2A.3.[16] Among those with exposure to high COVID-19 risk, the average alert times for cities in Clusters 4–6 is 10.8, 0.7 higher than that of those in Clusters 7–9. In contrast, according to the China City Statistical Yearbooks, the affected middle-commuting cities (Clusters 4–6), on average, have fewer employees in the tertiary sector, with a ratio of 57.5 in 2019, as opposed to 58.6 for those affected in Clusters 7–9. In addition, it is worth noting that a few cities exhibit essential travel patterns quite different from their neighbors. For instance, Lhasa, Ürümqi, Lanzhou, and Xining (provincial capitals of Tibet, Xinjiang, Gansu, and Qinghai) are in either the sixth or eighth cluster. As discussed, their overall intra-city

[16]We will discuss in the next subsection that exposure to high COVID-19 risk persistently weakened commuting behaviors.

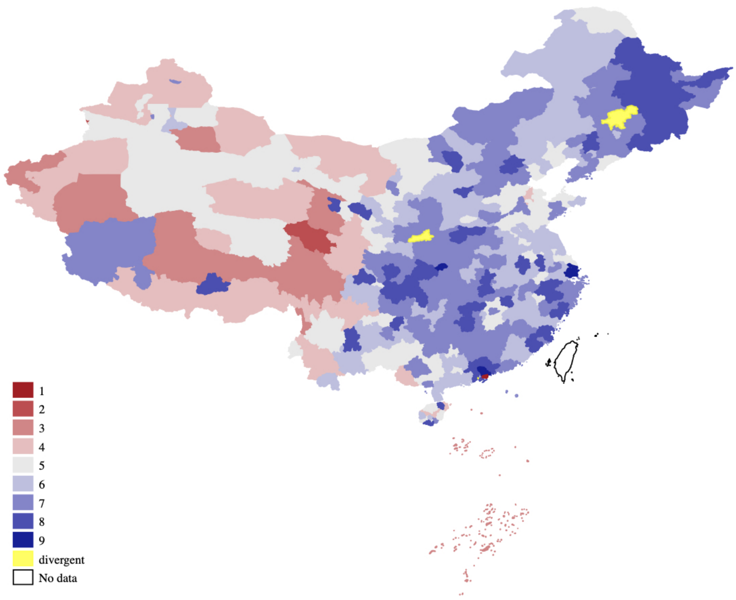
travel strength is less intensive than that of other cities within the same provinces and of other provincial capitals. In contrast, Hong Kong in Cluster 1 (see Table 2A.2), together with Macao, presents much higher commuting rates than other areas nearby.[17]

The patterns of the non-essential travel are flatter and more homogeneous. The transition paths of all the clusters plus three divergent cities are quite smooth, with the exception of Hong Kong, which falls increasingly behind. In contrast to essential travel, there is no significant division between regions. Further, Clusters 1–2 dominate across the country, with Lhasa, Ürümqi, Lanzhou, and Xining falling into Clusters 1–3. It bears emphasizing that increases in either the essential or non-essential travel are correlated with but do not necessarily cause decreases in the other. Hence, the lower mobility rates among western cities discussed previously should not be simply understood as higher percentages of non-essential trips.

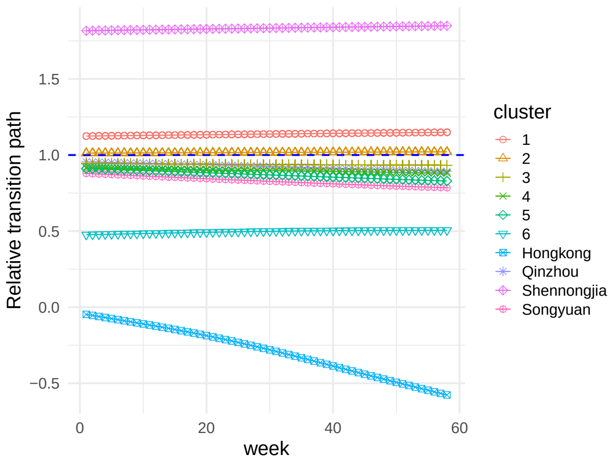
[17]Because data on HW and DLR travel are calculated as the proportion of each travel behavior, the population size of Baidu Maps users in Hong Kong and Macao does not matter here.



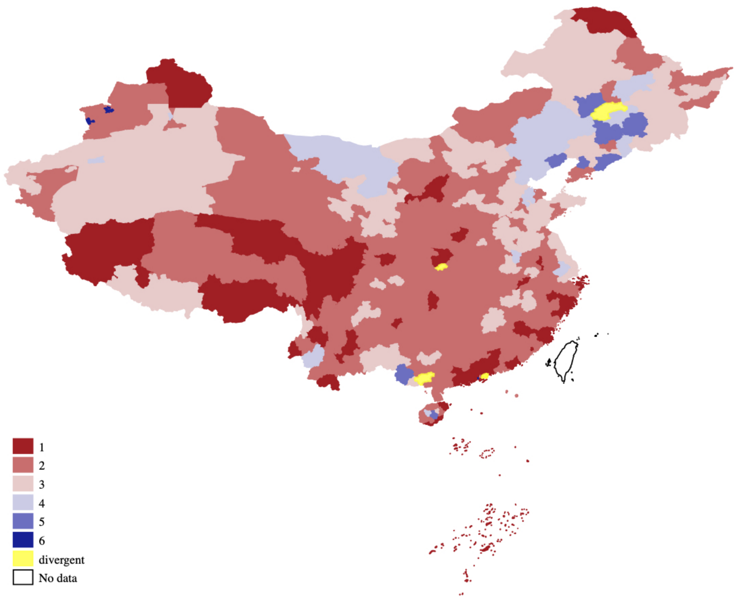
(a) Transition paths of HW commuting rates.



(b) Geographical distribution of HW clusters.



(c) Transition paths of DLR travel rates.



(d) Geographical distribution of DLR clusters.

Figure 2.2: Spatiotemporal results of HW and DLR travel during May 17, 2021–June 26, 2022.

The result of each city can be found in the Appendix and our dashboard.
Source: Created by authors using the Baidu Mobility Data.

2.4.3 Effects of exposure to high COVID-19 risk on commuting

Table 2.2: Aggregated treatment effect estimates

	(Pre-)treatment effects	(Post-)treatment effects
HW commute		
Aggregation types		
Simple weighted average	-0.0209*** (0.0033)	
Group-specific effects	-0.0204*** (0.0020)	
Event study	0.0002 (0.0001)	-0.0269*** (0.0033)
Calendar time effects	-0.0145*** (0.0032)	
DLR travel		
Aggregation types		
Simple weighted average	0.0073** (0.0031)	
Group-specific effects	0.0071*** (0.0011)	
Event study	-0.0004*** (0.0001)	0.0120*** (0.0029)
Calendar time effects	0.0034 (0.0031)	

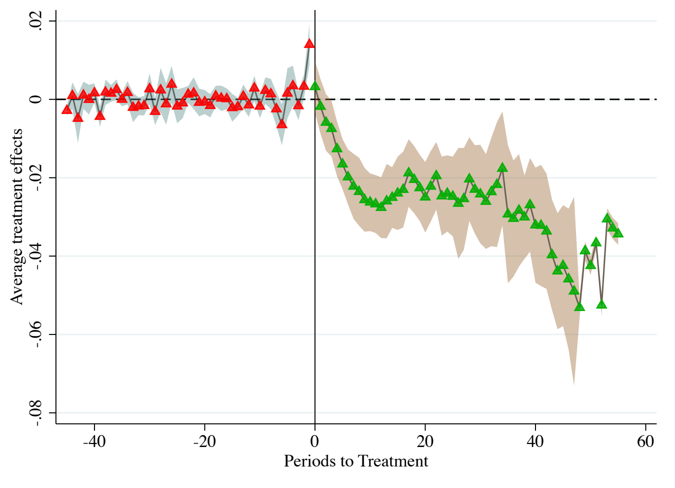
Notes: The row "Simple weighted average" presents the average treatment effects of all available groups across all periods. The rows "Group-specific effects" and "Event study" estimate average treatment effects based on the timing of any high-risk areas detected in the city for the first time during the studied sample period and the length of exposure to high COVID-19 risk, respectively. The row "Calendar time effects" reports average treatment effects of all available groups in each period. Robust and asymptotic standard errors are shown in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Source: Created by authors using the Baidu Mobility Data and COVID-19 risk-level data (Gong et al. 2023).

Table 2.2 reports the average treatment effects on essential and non-essential travel with four aggregation strategies. Foremost, the results of the first four rows are all statistically significant at the 1% level and indicate that once there were any high-risk areas in the city, commuting rates dropped accordingly by an average of 1.45%–2.69%. Given that we have 27 groups under the "Group-specific effects" aggregation, where the timing for cities to have high-risk areas varies, we also report each group's estimate in Figure 2A.4 where 25 out of 27 groups are found to experience negative treatment effects, with 23 of them being statistically significant at either the 1% or 5% level. In contrast, the next four rows indicate that exposure to high COVID-19 risk was correlated with an increase of 0.071%–1.2% in DLR travel rates. However, the interpretation for changes in DLR travel rates should hinge on HW commuting rates since being alerted to have high-risk areas by no means would encourage non-essential travel behaviors. Thus, the comparison suggests that high COVID-19 risk disproportionately influenced essential and non-essential travel.

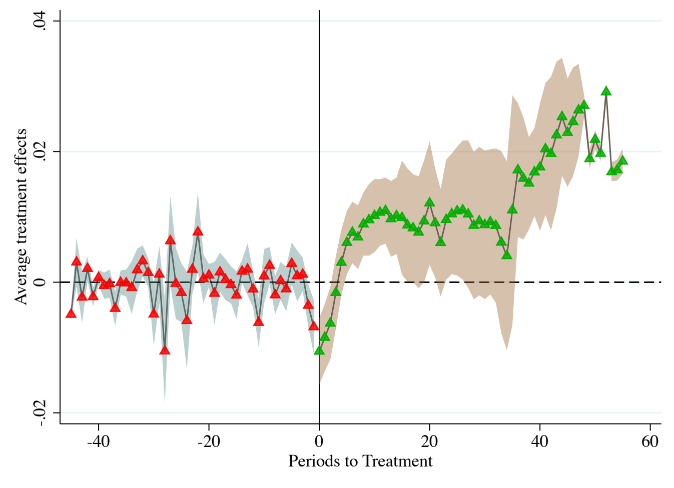
In addition, it is worth noting that the pre-treatment difference in HW commuting rates between treatment and control groups is statistically insignificant, with a coefficient smaller than 0.0002, in favor of the parallel trends assumption.[18] Similarly, the pre-treatment difference in DLR travel is statistically significant but only by a small magnitude of -0.04%, suggesting the pre-trends are relatively negligible. As illustrated in Figure 2.3(a), the declines in commuting rates of the treatment groups persisted throughout the post-treatment periods. In comparison, Figure 2.3(b) shows that the non-essential travel also decreased around the time of exposure to high COVID-19 risk but went up fast. As aforementioned, changes in non-essential travel partially mirror variations in the essential travel. The results thus suggest that although the emergence of high-risk areas and the implementation of containment measures influenced both travel behaviors, the essential travel was impacted more significantly and persistently likely due to the adaption to work-from-home.

Moreover, it is interesting to note that the estimates of the one to two periods right before the exposure (indicated by the black line in period 0) portray an increase (decrease) in the essential (non-essential) travel. It is plausible that as the Chinese

[18] Because weekly data are more sensitive to random fluctuations than quarterly or annual data, a few pre-treatment periods retain estimates distinguishable from zero but with a very small magnitude. While as observed in Figure 2.3(a), the pre-trends, in general, are not evident.



(a) Intra-city HW commute.



(b) Intra-city DLR travel.

Figure 2.3: Average treatment effects by the length of exposure to high COVID-19 risk in the city during May 17, 2021–June 26, 2022.

Notes: Estimates are visualized with 95% confidence intervals based on Callaway and Sant'Anna's (2021) DiD with multiple periods estimator and event study aggregation.
Source: Created by authors using the Baidu Mobility Data and COVID-19 risk-level data (Gong et al. 2023).

government reported the number of new confirmed cases every day, people made predictions about what might happen in the near future and reacted strategically, for instance, preparing for remote work and avoiding non-essential trips. In conclusion, the effects of detecting a high-risk area and introducing corresponding containment measures are far more profound than the minimum period documented for cities to move down to the low-risk level. This is probably because cities being alerted to have a high-risk area are more likely to experience high risk more often. As depicted in Figure 2A.3, none of the cities in the top three clusters had ever been exposed to high COVID-19 risk during the studied period, as opposed to the majority of cities with high-risk areas being alerted more than once. On the contrary, each cluster shown in Figure 2A.5 has some cities that were at high risk of COVID-19 contagion. Taken together, these findings provide further evidence that the impacts of high-risk shocks on essential travel were greater.

2.5 Conclusion

This chapter primarily assessed the intra-city mobility trajectories of 368 Chinese cities on a weekly basis during the transition from the Zero-COVID Policy towards reopening (Jan 17, 2022–Mar 12, 2023). It also examines the causal effects of exposure to high COVID-19 risk – which is a trigger for the highest-level restriction set by the risk-level system – on essential and non-essential travel behaviors during the strict containment period (May 17, 2021–June 26, 2022). Utilizing the latest Baidu Mobility Data and risk-level data, we find that the relative distance in intra-city mobility is decreasing within each cluster but widening between different clusters with cities, such as Lanzhou, Xining, Lhasa, and Ürümqi, that lag behind already increasingly falling behind. A clear division is found between Western China and the rest of the country in the cases of both intra-city travel strength and HW commutes. Further, the alert of and intervention in high-risk areas persistently had stronger effects on commuting rates than on non-essential travel rates. In sum, this study contributes to the literature on spatiotemporal mobility and COVID-induced socio-economic consequences and is the first applying the sophisticated clustering technique proposed by (Phillips and Sul 2007, 2009) to assess short-term variations on a weekly basis. To our knowledge, no other method to date is able to study co-movements of series from a statistical perspective. As it enables us to measure both the speed and degree of the recovery of daily travel, which is usually within cities, it also suits the scenario of reopening well.

Our novel findings add to the emerging discussions on new mobility patterns in the (post-)pandemic era, particularly providing potential explanations beyond benchmark data for the contrasting results observed in Liu et al. (2021) and Mu et al. (2023). With an increasing number of cities identified to have high-risk areas, cities that are more susceptible to reduced commutes will exhibit lower levels of local travel strength, and if, by the end of the respective studied periods, the cohort of affected cities coincidentally tends to be less or more economically developed than the unaffected, the results will be different. Further, our results complement the existing evidence for COVID-related spatial and economic disparities (e.g., Chakraborty and Mukherjee 2023; Spelta and Pagnottoni 2021) by uncovering multifaceted factors behind the widening gaps in intra-city mobility between clusters: first, the initial mobility levels of lower clusters tend to be lower, second, lower clusters have experienced the most significant decreases during the turbulent period (Aug 2022–Dec 2022), and third, the recovery rates of lower clusters tend to be slower. In other words, both lower pre-pandemic mobility levels and COVID-related influences have jointly dragged these clusters to be further behind. To conclude, this study highlights two key points. Firstly, it demonstrates that implementing one-size-fits-all policies is not ideal in the context of post-pandemic travel policy-making. Instead, a case-by-case approach at the cluster level is recommended, allowing policies to be tailored to the specific needs of each cluster. This is particularly important for the lowest two clusters, mainly comprised of cities in Xinjiang and Tibet. Secondly, our findings emphasize the need for supportive policies devised in a progressive manner, such as discounts for local transportation and subsidies for the vulnerable population to promote mobility and consumption, as simply lifting or removing stringent containment measures is not sufficient. By doing so, elementary remedies can be promptly introduced to achieve a better trade-off between health and economic performance.

Chapter 3

Valuing Children in China: Parents' Perceptions, Spending Priorities, and Children's Capabilities

3.1 Introduction

Debates on children's capabilities center on the essence of human development and flourishing lives of humans. Such debates often focus on the instrumental role of education systems (Hanushek and Woessmann 2008; Hanushek et al. 2016; World Bank Group 2018), early environments (Cunha and Heckman 2007; Heckman 2008a), parental practices (Liu et al. 2020a; Vasilyeva et al. 2018), children's health (Goldhagen et al. 2020; Gunnar et al. 2020), and children's nutrition (Black et al. 2020; Shrestha et al. 2021), among others, in developing their capabilities. Regardless of how big the family is or how constituted, families are predominantly responsible for making necessary arrangements to develop their children's capabilities (Nussbaum 2000). Particular types of distinctive good, such as a "relationship good", only become possible because of the existence of families (Swift and Brighouse 2014). Families are important because of the investments they make in their children; not in the sense discussed by Becker (1974) and Becker and Tomes (1976), where children are regarded as competition in families' consumption decisions, but rather from the perspective of human development and the capability approach (Nussbaum 2000, 2011; Sen 1997, 2009, 2017).

Within this context, children's development depends on their capabilities that parents have "reason to value" (Sen 1999), that is, the valuable capabilities that parents wish to foster in their children. Despite this, parents' priorities have rarely been examined, and only indirectly estimated by children's academic studies, or simply ignored (Biggeri and Mehrotra 2011). Consequently, these types of studies prevalently assume that parents in general attach a similar level of importance to their children's development. They ignore the fact that families' prioritization of spending in relation to their children could reflect a concerted judgment about the value of education and their children's future, which, in turn, affects their prioritization.

This approach applies to some recent studies examining multidimensional poverty (Zhang et al. 2021) and child development in China (Chen et al. 2021; Cui et al. 2019; Sylvia et al. 2022) that focuses on the impact of parenting interventions. However, if family resources are accorded importance based on the use to which they are put, and if the use depends on how different capabilities are valued, it is essential to assess how families prioritize their resources. Given this background, the main objective of this chapter is to examine how the academic performance of Chinese children and adolescents depends on household spending priorities vis-à-vis parental practices and other factors. To address these issues, several new indicators have been developed, such as the "parent advantage index" (PAI, modelled on the Human Development Index) and the "spending priority ranking" (SPR, based on a ranking of ten spending category groups assessed as a proportion of a family's household expenditure). Further, children's academic development is measured through an assessment of their learning outcomes and learning processes.

The nationally-representative China Family Panel Studies (CFPS) surveys provide a rich database of useful variables, such as school quality, parental reactions to children's unsatisfactory test scores, and children's study habits and discipline (SHD). The sample used for this study comprises 8,422 Chinese children and adolescents surveyed during 2012 to 2018. Our study aims to contribute to the literature on children's development by assessing the place of families in children's academic performance (Heckman and Mosso 2014), the relevance of traditional Chinese culture that cultivates positive attitudes towards learning (Hsu and Wu 2015), and the significance of parental prioritization of categories of spending, as an expression of attitudes and values (Nussbaum 2000; Sen 1997, 2017). The study does not simply consider the impact of family resources or parental practices on their children's development, but extends to how families value education, particularly through spending prioritization.

Overall, our results confirm the effects of parental advantages and a higher spending priority on educational and cultural activities, among others, on children's academic development. The results proved robust on a series of alternative estimations, including Lewbel's (2012) heteroskedasticity-based instrumental variables.

3.2 Theoretical Framework: Children and the Capability Approach

There are many different approaches to assessing children's human development, the most-frequently used being the Capability Approach (Nussbaum 2011; Sen 2017; Yousefzadeh et al. 2019). Despite analytical nuances, a common feature of this approach is its emphasis on informational pluralism. In essence, pluralism is argued for evaluations based on information-rich accounts of the state of affairs. It is not simply a matter of elaborating multidimensional indicators, but is also concerned with using different informational spaces to construct these indicators. The most widely-used informational spaces in the literature of impact evaluation are: resources, subjective well-being, rights and capabilities (Comim 2021; Sen 1980, 2017). Sen's argument for "the impossibility of a Paretian liberal" was an important milestone in this debate, providing as it does a compelling formulation of rights in the literature of welfare economics (Suzumura 2011).

The pluralist nature of the Capability Approach implies a valuational exercise that demands, firstly, a consideration of a multiplicity of dimensions and variables in the relevant spaces (that are more often than not heterogeneous), secondly, a solution to the complex issue of the varying importance of different functioning and capabilities and how they are evaluated, and thirdly, an engagement with the "agency aspect" that the approach highlights. In fact, the elaboration of information-rich accounts is not the only important element in the Capability Approach, because these spaces should automatically be part of accounts that attempt to explain how autonomous actions reflect, in different degrees, a person's freedom to live in a way that they would value

(Sen 1999).

Autonomous actions are difficult to characterize. Nonetheless, the Capability Approach supports assessments grounded in reasoned scrutiny as a way of capturing agency (Sen 2009). This means that information should incorporate individual and collective reflective evaluations about what people have reasons to value. In theory, reasoned scrutiny represents a strong critique against assessments based on mechanical judgments. While, in practice, operationalizing the scrutiny is often challenging once the exercises of prioritization attached to the selection of key capabilities are few. Another key feature of the Capability Approach, particularly in Sen's formulation, is a conceptual distinction between comprehensive outcomes (those that include the processes of choice) and culmination outcomes (those that only display the final results of the act of choice). This is because the act of choice also has process significance within which results should be characterized, not only for the final results, but also for all those features of the processes that final results involve (Sen 2002). Thus, different results obtained from different processes cannot receive the same evaluation. Because the Capability Approach values individual autonomy, as discussed below, it is not sufficient to be concerned only with what an individual receives should they choose, but that they actually get to choose what they receive themselves. Thus, whenever we examine children's outcomes, we are concerned not only with the marks from their exams, but also with the learning processes involved in achieving those results.

The use of the Capability Approach for assessing children's capabilities invites us to look at childhood from a different perspective, integrating key aspects of:

1. the role of families in promoting human development, focusing on how children are raised,

2. the path and time-dependent dimension of practices and policies that foster children's development,
3. the recognition of children's agency and autonomy, and
4. the role of emotions during childhood.

Families are important in promoting children's capabilities, as networks of love and care(Nussbaum 2006). It is within families that children grow up to become fully-functioning human beings and where they learn to become moral agents, particularly in early childhood. Parental practices can often be categorized as distinct parenting styles that portray certain behavioral and attitudinal patterns towards children. The most influential styles are:

- authoritative: evident when parents show understanding, open communication, respect and emotional support and considered the most effective and loving parenting style,
- authoritarian: evident when parents rigorously assess their children's behavior, impose rigid norms and punishment without sympathy for children's difficulties, typically exemplified by an absence of emotional support,
- permissive: evident when parents acquiesce in actions and behaviors as their children please, still probably being loving and sympathetic yet not responding with discipline and control. This is often the case with absent parents who try to compensate for their absence by indulging their children, and
- negligent: evident when parents do not show much interest in their children's development. Their involvement is minimum, with parents spending little time with their children and offering little or no level of support and control.

Different parenting styles entail different prioritization strategies that parents use to manage their children's human development. A variety of circumstances can embody the prioritization, namely, by the amount parents spend on their children's education, the time they dedicate to play with their children, whether they help with homework or not, and whether parents support their children emotionally. While parenting styles might not be empirically evident in the clear-cut descriptions above, this categorization is useful in calling attention to the diversity of processes of raising children.

A common mistake in assessing children's development is viewing childhood as a single discrete period in one's life, without considering different stages of child development as being unevenly affected by biological and neurological factors (Borghans et al. 2008; Cunha and Heckman 2007). Indeed, children's receptiveness to language learning is higher by 3 years of age, their IQ scores are often stable by the age of 10, and emotions and self-regulation from the malleability of the pre-frontal cortex lasts until the end of adolescence (Heckman 2008b; Rose and Fischer 2011). Correspondingly, investment in early childhood education should be distinguished from that in late childhood, and the impact of parental investment on children's skills and human development also depend on sensitive (more effective) and critical (unique) periods. Time is of the essence in the matter of being a child, given how their development is uniquely sensitive to different flows of time and timing of particular interventions. Time also matters from another angle – children need time with their parents, time for playing, time for being creative, and time in which they are protected and can flourish.

The use of the Capability Approach also encourages consideration of how children develop their own capabilities. As much as children require some basic functionings, such as compulsory education, before they can fully exercise their autonomy (Nussbaum

2011), it is important to acknowledge that children have a certain capacity for self-determination that is exercised from a very early age (Ballet et al. 2011; Saito 2003). Children will have full-fledged autonomy when they become adults, with conceptions of what is right and what is good developed with faculties enabled during childhood (Rawls 1971). Whereas a paternalistic view sees children as vulnerable and dependent on their parents, a capability view sees children as an evolving project of human self-determination. The debate is not that children are unable to make choices, but that they may not be able to evaluate and revise the choices they make. Evidence suggests that children start to learn to be independent of their parents from an early age (Lansdown 2004). No one claims that children can display the same level of self-determination as do adults. The flaw is in denying to children a capacity for self-determination that is progressively evolving. In fact, children can persuade adults of what they want, and they can negotiate and renegotiate boundaries imposed by adults (Alderson 2001; Anich et al. 2011; Punch 2002). This means that a child's human development should not be seen merely as a result of his or her parents' priorities, but rather as an interactive process between parents and the expression of the child's own agency (Bellanca et al. 2011).

Finally, we refer to the role of emotions in shaping children's ethical reasoning (Nussbaum 2011; Nussbaum 2006). While emotions help to explain children's motivation for acting and their endurance (Biggeri et al. 2011b), they also have an important cognitive role. As Nussbaum (2011) and Cunha and Heckman (2007) demonstrate, emotions can be decisive for the formation of children's deliberative beliefs, enabling them to perceive critical features in a situation. An example is useful here. A father singing nursery songs to his baby daughter fosters the baby's moral life and, as such, can be understood as a key practice to be respected and supported in the promotion of the child's future capabilities. The human sense of value is built upon such interactions within which emotional cognition plays an important role.

Assessing children's human development through a capability lens means that we should look at it from a multidimensional and pluralist perspective, analyzing how families define their priorities about what they have reason to value related to their children's development. Further, it means going beyond the concept of children as beings without emotions or will. Understanding the formation of children's capabilities, therefore, entails seeing them as quintessentially dynamic and time-dependent. Of course, it is difficult to emphasize all these different elements in a single analytical discussion. For this reason, we highlight here the links between parental practices and spending priorities, and the impact they might have on children's cognitive development.

3.3 Data Description and Variable Definitions

3.3.1 Data Source

We based our empirical analysis of children's capabilities on the China Family Panel Studies (CFPS), a nationally-representative survey launched in 2010 by the Institute of Social Science Survey, Peking University. The CFPS was designed to collect data biennially at the individual, household, and community level from 25 provinces, municipalities, and autonomous regions, representing 95% of China's population. Information on children and adolescents within surveyed households was separately collected.[1] The attrition rate is around 25% biennially. In this study, 4 out of 5 waves were merged to as-

[1] All questionnaires were filled in by parents on behalf of their children under 10 years old at the time of the survey.

semble a panel covering data from 2012 to 2018, with child-, adult- and household-level data matched year-to-year. The 2010 survey data was not included because variables provided in the later waves were absent in this initial version. Therefore, our panel includes 8,422 children and adolescents aged 6 to 16 who attended school.

3.3.2 Child Academic Development Index

Children's capabilities comprise a rich array of cognitive and socio-emotional dimensions. If we were to follow Nussbaum (2018) list of central capabilities, to construct a comprehensive index, we would need indicators of children's senses, imagination and thought, emotions, practical reasoning, and sense of affiliation or even of play, including the ability to laugh. Unfortunately, data for this kind of evaluation are not normally found empirically. For this reason, we adopted a modest approach, concentrating on a core aspect of children's human development, namely, their academic development. However, our indicator goes beyond just test scores in literacy and numeracy to include children's study habits and other non-cognitive elements.

It is important to note that, in the Chinese education system, test scores are decisive information for children to move forward in the education system and the main mechanism for entry into prestigious schools. Consequently, the Gaokao (the national college entrance examination in China) is deemed a major turning point in the life course that determines a person's career opportunities, earning potential, and even marriage prospects. Studying for the three-day Gaokao can be likened to a marathon in which a variety of cognitive, non-cognitive, and environmental factors connect to shape the outcome. Following Amartya Sen's distinction between culmination and comprehensive outcomes (Sen 2002), we focus on children's academic development by structuring it into two parts: one examining learning outcomes (corresponding to culmination outcomes) and the other taking into account learning processes (to reflect comprehensive outcomes).[2]

The CFPS provides two ordinal variables scored by parents, describing children's academic achievement in Mathematics and Chinese based on their performance in the previous semester. We calculated the average point of the two subjects and normalized the result. Likewise, learning processes were measured as the normalized average of seven questions evaluated by parents regarding how good their child's study habits and discipline (SHD) were.[3] Questions are summarized in Table 3A.1. The children's academic development index was formulated as follows:[4]

$$\begin{aligned} Score &= \frac{Mathematics + Chinese}{2} \\ SHD &= \frac{QA1 + QA2 + ... + QA7}{7} \\ Norm\_Score &= \frac{Score - Min(Score)}{Max(Score) - Min(Score)} \\ Norm\_SHD &= \frac{SHD - Min(SHD)}{Max(SHD) - Min(SHD)} \end{aligned} \tag{3.1}$$

$$CADI = \frac{Norm\_Score + Norm\_SHD}{2} \tag{3.2}$$

It is important to note that parents' understanding of their children's capabilities might be more relevant than the capabilities *per se* in this case, given that test scores are always limited in what they test, while parents can have a more comprehensive knowledge of their children's skills informed by sequences of tests and other indicators that naturally

[2]In addition, Heckman et al. (2006) pointed out that aside from cognitive abilities, socio-emotional skills, such as perseverance, motivation, and self-control, had a direct impact on schooling decisions and test scores. Cunha and Heckman (2007) also demonstrated the interplay of cognitive and non-cognitive skill accumulation.

[3]The pairwise correlation rate of the seven dimensions of SHD is between 0.22 and 0.50.

[4]Despite the dimensional distinction, test scores and SHD are both an indication of academic development, a latent construct. With this in mind, we also ran a common factor analysis (CFA) to generate an alternate CADI, which accounts for 90.1% of the common variance of the nine variables.

enter into the formation of their views. They are also key observers of their children's study habits that incorporate a wide range of non-cognitive elements. Moreover, both methods and difficulties of tests varying substantially across schools and regions could also introduce considerable noises to using objective scores as the evaluative information of student achievements.

3.3.3 Parent Advantage Index

Socio-cultural factors, particularly family-related, influence the formation and development of children's mindsets and behaviors. Thus, we created three variables to account for parents' beings and doings. Among them, the parent advantage index is a unique proxy for functionings. As aforementioned, the PAI is based on the Human Development Index using proxies for the HDI dimensions, such as self-evaluated health status, years of schooling and the natural logarithm of household income per capita.[5] We adopted household income per capita to account for intergenerational financial transfers, which commonly occur in Chinese families, that may be driven by the traditional culture of filial piety (see e.g., Sun 2004; Zhu 2016).[6] Furthermore, previous studies pointed out that material (housing or financial) support and living arrangements were often intertwined (see e.g., Li and Wu 2019; Yi et al. 2018; Yin 2010). In this sense, the benefits or burdens placed on adults by their elderly parents need to be considered when assessing their economic advantages.[7]

[5]The function converting from additional income to enhanced capabilities is likely to be concave (Anand and Sen 2000), so the natural logarithm of income is often used in the HDI. A body of literature also demonstrates the non-linearity of family income and child outcomes, see Cooper and Stewart (2021).

[6]"Filial piety" refers to Chinese ethics rooted in Confucianism, emphasising attitudes of obedience, respect, care, and love towards parents.

[7]In rural China, adult children are more likely to provide financial support to their elderly parents than their urban counterparts (Lee and Xiao 1998). In contrast, due in a large part to sky-rocketing housing prices, adult children living in urban locations are now more likely to receive financial support from their elderly parents (Rosenzweig and Zhang 2014). Further, Silverstein and Zhang (2020) found that grandparents tended to provide economic resources to their grandchildren in rural China.

The PAI follows the pluralist approach advocated by Sen and Nussbaum once it combines a subjective variable (self-evaluated health status) with a variable of functionings (years of schooling) and another grounded in resources (household income). It does not differ significantly from the traditional HDI in conception and, as such, cannot be interpreted as an index of capabilities, although it follows a key characteristic of the capability approach.

3.3.4 Spending Priority Ranking

Several methods are used in the capability literature to select relevant capability information (Burchardt and Vizard 2011; Byskov 2018). However, these methods often focus on how researchers can identify basic or key capabilities from statistical data without directly tackling the reasoned scrutiny aspect of capabilities, where individuals' priorities are represented by particular indicators. As much as this is not a trivial task, being able to signal how individuals translate their "reasons to value" into specific priorities is essential under this framework. For this reason, we have built an indicator that attempts to represent individuals' priorities through their budget allocation choices. We classified 26 expenditure items into ten groups, according to Xie et al. (2017), and calculated spending on items within each group. By calculating the ratio of spending in each group to total expenditure, we then ranked the groups, associating higher percentages of spending with higher levels of priority.[8] In other words, the number "10" represents the highest priority here. Previously, Ratigan (2017) compared social policy priorities of Chinese provinces in a fashion similar to ours.

As shown in Figure 3.1, nearly 47% of households gave the highest priority to the

[8]The pairwise correlation rate of the ten dimensions of the SPR is between -0.01 and -0.22.

"diet-relevant" spending group, while the largest share of spending on the transport-relevant group was 21%, ranking eighth. Similarly, for the group of "rent and utilities", a sixth-level priority given by around 17% of households was the most common, and the prevailing rank seen in the "necessity-relevant" group was ranked fifth. Further, medication, healthcare, and sports items were prioritized fourth by approximately 20% of the families. In contrast, the "education-relevant" group was more significant, ranked ninth and tenth by 30% of the families in total. This is the key variable of interest in our analysis since it reflects how households value investing in their children. While spending on clothing-relevant items was most often ranked sixth and fifth. In addition, the remainder of the spending groups had similar distributions, with the donation-, insurance- and all-other-relevant consumption ranked third by 46%, 42%, and 40% of the households, respectively. All these variables are further summarized in Table 3.1.

We also created a group of supplementary SPR indicators in the same way by adding housing mortgage to the utility-relevant group and re-ranked spending groups accordingly. Corresponding statistics are reported in Table 3A.2. However, only 73% observations are retained because the CFPS didn't survey households' housing mortgage in 2012.

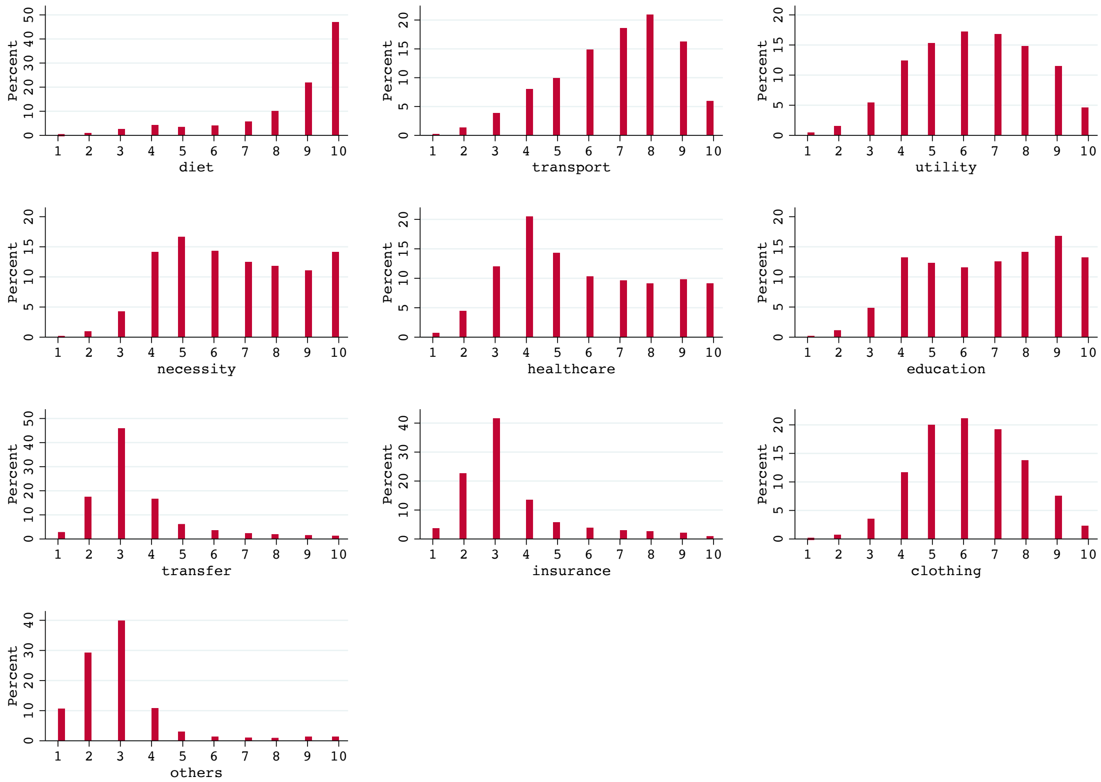


Figure 3.1: Histograms of the SPR indicators.

Notes: By design, larger numbers here reflect higher priorities.
Source: Authors' elaboration using CFPS data.

Table 3.1: Descriptive Statistics – Spending Priority Ranking

Variable	Description	Obs.	Mean	Std. Dev.	Min	Max
Spending Priority Ranking						
diet	Spending priority for food and drink including eating out	17,291	7.94	2.32	1	10
transport	Spending priority for local transportation and post, and telecommunications	17,291	7.03	1.97	1	10
utility	Spending priority for rent, utilities and property management	17,291	6.48	2.02	1	10
necessities	Spending priority for daily necessities, home repairs, cars, other transport tools, furnitures and electrical appliances	17,291	6.56	2.20	1	10
healthcare	Spending priority for medication, healthcare and fitness	17,291	5.76	2.39	1	10
education	Spending priority for education, culture and recreation, and travel	17,291	6.95	2.24	1	10
donation	Spending priority for financial support given to others and social donation	17,291	3.73	1.79	1	10
insurance	Spending priority for business insurance	17,291	3.36	1.73	1	10
clothing	Spending priority for clothes and beauty (e.g., haircut, spa, cosmetics)	17,291	6.40	1.73	1	10
other	Spending priority for all other items	17,291	3.29	1.95	1	10
Time						
year	Year of survey	17,291	2013.33	1.47	2012	2019

Notes: The amount paid for some items was asked on a monthly basis. For calculation purposes, we converted them into annual quantities. The larger the value, the higher the priority. All statistics were adjusted using the sampling weights.
Source: Authors' elaboration using CFPS data.

3.3.5 Other Covariates

In addition to the CADI, PAI and SPR variables, we introduced 11 covariates to control for heterogeneity across samples; five at the child level and six at the parent and household level. These variables were derived from the stories that the CFPS data illustrate and assessed from a capability perspective. We elaborate on both these levels here.

Child-Level Covariates

The CFPS collected rich information concerning children's development. For instance, it records how many times a child went to hospitals or clinics in cities, towns, communities, or villages in the last 12 months. One can infer the status of children's health from this information. In addition, the CFPS registered information about whom the actual carers of a child were, and how many times per week, on average, the child met at least one of his or her parents. Based on these questions, we created a dummy specifying children who were not looked after by their parents in person and who saw neither of their parents per week.[9] Furthermore, the quality of teaching, the school atmosphere, and the peer culture can be associated with children's academic development (Lynch et al. 2013). In the context of China, attending a key school is both a reflection and a determinant of student achievement.[10] Hence, we created another dummy to distinguish students in key schools from those in ordinary schools. We also included gender and sleep time to control for individual heterogeneity.

Parent- and Household-Level Covariates

A substantial body of literature examined the effects of parenting style on child health and student achievement (e.g., Burton et al. 2002; Cui et al. 2019; Dooley and Stewart 2007). Two variables were created in our dataset to control for parents' functionings. One accounts for parental practices, measuring the extent to which parents paid attention to and monitored their child's learning and recreational activities.[11] We

[9]Questions used to justify this factor are listed as QC1–QC3 in Table 3A.1.

[10]Chinese secondary schools are divided into "key" and "ordinary" schools. Designated key schools distinguish themselves from ordinary schools by their academic reputation and generally gain more educational resources in areas such as teachers, equipment, and funding. Students need to compete for admission to key schools, meaning that only the best cohort is entitled to study there.

[11]The pairwise correlation rate of the six dimensions of parental practices provided by parents themselves is between 0.16 and 0.39.

also incorporated information based on the observations of home environment by the CFPS's interviewers, assessing to what extent parents were concerned with their child's education and actively communicated with their child. Questions in this regard are summarized in Table 3.1 (QB1–QB8).

A further covariate accounts for parents' reactions to their children's unsatisfactory test scores, which is closely related to our dependent variable (DV). The possible reactions are summarized as: (a) contact the teacher; (b) physical punishment; (c) scold the child; (d) ask the child to study harder; (e) ground the child; (f) help the child more; (g) take no reaction. We divided these options into three categories, i.e., negative, passive and positive.[12] Further, other covariates are the average age of parents and several dummies that refer to parental marital status, household residence (in urban or rural area), and a family's savings for children's education, respectively. Descriptive statistics of other variables are given in Table 3.2.

[12]Reaction (b), (c) and (d) are classified as negative. In contrast, reaction (a), (d) and (f) are classified as positive. Only reaction (g) is considered passive.

Table 3.2: Descriptive Statistics – Children, Parents & Households

Variable	Description	Obs.	Mean	Std. Dev.	Min	Max
Child Level						
CADI	Normalized arithmetic mean of learning processes and outcomes	12,658	0.6220	0.1841	0	1
score	Test scores	14,433	2.78	0.0298	1	4
SHD	Study habits and discipline	14,270	3.57	0.0182	1	5
child gender	Boy=1; Girl=0	17,291	0.5144	0.4998	0	1
child age	Age	17,287	10.83	2.97	6	16
key_school	Enrolled in a key school=1; Not=0	13,764	0.2514	0.4338	0	1
sleep	Hours of sleep on weekdays	16,520	9.03	1.02	5	13
child health	Frequency of visiting hospitals and clinics in the last 12 months	16,987	1.12	2.37	0	122
absence	Not staying with and rarely see parents=1; Otherwise=0	17,038	0.0573	0.2324	0	1
Parent/Household Level						
PAI	Normalized arithmetic mean of health, education and household income per capita	13,875	0.4117	0.1056	0	1
health	Self-evaluated health status	32,441	3.13	0.0325	1	5
education	Years of schooling	32,022	2.70	0.0906	0	18
family_income	Household income per capita (log)	31,108	8.80	0.0603	-1.61	15.23
parent_age	Average age of parents	17,134	38.32	5.46	22	83
marital	Divorced or widowed=1; Married=0	14,725	0.0429	0.2027	0	1
parental practices	Normalized geometric mean of parental practices	12,403	0.6170	0.1482	0	1
reaction	Reaction to child's unsatisfactory test scores: positive=3; passive=2; negative=1	15,613	2.78	0.6108	1	3
edu_savings	Saved money for child's education=1; Not save=0	16,965	0.2043	0.4032	0	1
child number	The number of children	17,291	1.67	0.8478	1	8
residence	Live in urban areas=1; Rural=0	16,836	0.5220	0.4995	0	1

Notes: All statistics were adjusted using the sampling weights.
Source: Authors' elaboration using CFPS data.

3.4 Econometric Modelling and Techniques

We undertook a threefold empirical analysis. To alleviate potential endogeneity, we firstly adopted a fixed-effects model. Since some informative variables are time-invariant and our core predictors are slowly-varying household-level variables,[13] a within individual or household estimator is not applicable. Yet all individuals were nested within their households, we thus checked the robustness of the results by considering multilevel modeling, which helped eliminate household-level unobserved heterogeneity. Given that our data contain 6,405 households and 8,422 children, variations estimated in this way were within-child for the majority of the observations, acting as an ideal alternative to individual fixed effects. Finally, we double-checked our analysis with Lewbel's (2012) heteroskedasticity-based instruments. This approach is widely used when external instruments are not available (e.g., Chung et al. 2020; Liu and Yu 2020).

3.4.1 Fixed Effects Models

The main specification is as follows:[14]

$$CADI_{it} = \alpha + \beta_1 PAI_{it} + \beta_2 SPR_{it} + \beta_3 X_{it} + \gamma_j + \gamma_B + \gamma_N + \gamma_t + \varepsilon_{it} \tag{3.3}$$

where $CADI_{it}$ is the child academic development index of a child i living in province j identified in wave t. PAI_{it} is the average parent advantage index of child i's parents; SPR_{it} are spending priorities of child i's household;[15] X_{it} is a row vector of individual

[13] 3476 children were surveyed only once. In addition, half of the rest observations have a difference of numeric value no more than 1 in spending priorities between survey waves.

[14] Additionally, we estimated this equation using the CFA-based CADI and two components of the CADI, i.e., test scores and SHD, and report the results in the Appendix.

[15] We also estimate supplementary SPR indicators to further check the results.

and household characteristics. γ_j are province fixed effects; γ_B are child i's birth year fixed effects; γ_N are the number of children fixed effects; γ_t are the survey wave fixed effects. ε_{it} is an idiosyncratic error term.

The survey wave fixed-effects allow us to avoid systematic differences imposed by time across the four waves used in this study. The province fixed-effects help eliminate unobserved macro factors existing at the provincial level, such as differences in educational standards and Gaokao policies. Further, the birth year fixed-effects reduce two concerns. Firstly, new education-relevant policies could be issued every few years, altering parents' attitudes and children's learning experiences. Secondly, as shown in Figure 3.2, one component of the CADI, test scores, tends to decrease for children who become older. Additionally, as China had widely implemented the One-Child Policy for over two decades until late 2015, some invisible but ingrained differences in the families, such as a strong son preference, a lack of contraception or abortion, and religious factors, would be expected between one-child and multiple-child families. Even within multiple-child families, the number of children would most likely have an essential impact on household spending behaviors and the value placed on each expenditure item.

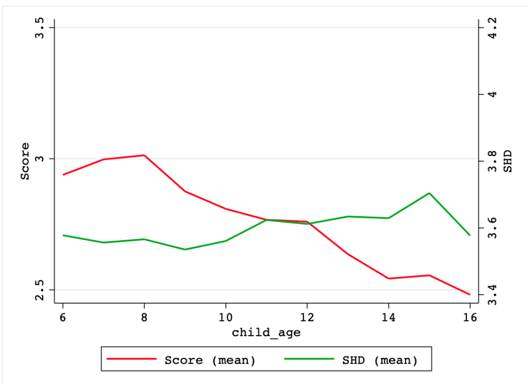


Figure 3.2: Test scores and study habits and discipline against child age.

Source: Authors' calculation using CFPS data.

Beyond the priority ranking, it is of interest to know if conversion factors in our model form differential effects on child outcomes. For instance, for children who attended a key school, it is likely that educational resources were converted to learning outcomes more efficiently than for their ordinary-school counterparts, *ceteris paribus*. Following Comim et al. (2018),[16] we estimate conversion rates by including interaction effects between four dummies and the SPR indicators.[17]

Therefore, we added the interaction terms, one by one, to the model, as follows:

$$CADI_{it} = \alpha + \beta_1 PAI_{it} + \beta_2 SPR_{it} + \beta_3 X_{it} + \beta_4 (SPR_{it} * D_{it}) + \beta_5 D_{it} + \gamma_j + \gamma_B + \gamma_N + \gamma_t + \varepsilon_{it} \tag{3.4}$$

where D_{it} is the dummy variable accounting for the conversion factor of interest. Other variables remain the same, as above.

3.4.2 Multilevel Modeling

Members of the same family tend to be similar owing to their presence in the same household and their common upbringing. This issue can be suitably tackled by multilevel modelling which is often used to estimate hierarchical and clustered data (e.g., Arunachalam et al. 2020; Krumbiegel et al. 2018; Liebenehm 2018; Smith et al. 2017; Yergeau 2020). Besides accounting for between-household heterogeneity correlated with both the DV and explanatory variables, we also treated provinces where households were nested as an additional level, higher than the household, to absorb unobserved province-specific effects.

[16]This approach can avoid involving a non-parametric first stage, as do non-linear or frontier models, where functioning and resources cannot be clearly distinguished (e.g., Binder and Broekel 2011, 2012a,b). Further, compared to a sub-group analysis, it more precisely estimates the differential effect of the resources for each category of conversion factors.

[17]The dummies are gender, tertiary education of parents (equals 1 if either of the parents received higher education, and 0 otherwise), urban area, and key school.

Here, let i,k and j represent the level-1 (individual), level-2 (household) and level-3 (province) unit, respectively. A three-level hierarchical linear model can be defined as follows:

$$
\begin{aligned}
CADI_{it} &= \alpha_{0jk} + \beta_1 PAI_{it} + \beta_2 SPR_{it} + \beta_3 X_{it} + \gamma_t + \varepsilon_{ijkt} \\
&\text{where } \alpha_{0jk} = \mu_{000} + \mu_{00j} + \mu_{0jk}
\end{aligned}
\tag{3.5}
$$

where μ_{000} is the mean intercept, common to all observations; μ_{00j} is the between-province error term, reflecting the deviation of province j's mean from the grand mean of the provinces; μ_{0jk} is the between-household error term, reflecting the deviation of household k's mean from the grand mean of the households within province j; $\varepsilon_{ijkt}$is the between-individual error term, absorbing the difference between the CADI of child i in household k and household k's average CADI. Definitions of other variables remain the same, as above.

3.4.3 Two-Stage Least Squares Models Using Lewbel's Instruments

$$
\begin{aligned}
&\text{1st stage: } SPR_{it} = X'\beta_1 + \varepsilon_1 \\
&\text{2nd stage: } CADI_{it} = X'\beta_2 + SPR_{it}\gamma_1 + \varepsilon_2
\end{aligned}
\tag{3.6}
$$

where X is a vector of exogenous covariates; SPR_{it}, our core predictors, are endogenous variables that need constructing valid instruments; ε_1 and ε_2 are error terms. Suppose Z is an element of X. An exclusion restriction of standard instrumental variables (IV) estimation assumes $Cov(Z, \varepsilon_2) = 0$ and $\beta_1 \neq 0, \beta_2 = 0$ for $X = Z$.

In the case of no instrument available, the 2SLS model cannot be identified in the usual way, while Lewbel (2012) proves that with some heteroscedasticity of ε_1, identification can be obtained by having two key assumptions: $Cov(Z, \varepsilon_1 \varepsilon_2) = 0$ and

$Cov(Z, \varepsilon_1^2) \neq 0$, where Z can be a subset of the elements of X. $Cov(Z, \varepsilon_1^2) \neq 0$ can be empirically tested by applying a Pagan and Hall (1983) test and Hansen's (1982) J-test can be implemented to check validity of the method if the model is overidentified (Baum and Lewbel 2019).[18]

3.5 Main Results

In Table 3.3 below, we report the results of five models: a baseline model where no covariates but the PAI and SPR indicators are estimated; a FE model where all covariates and fixed-effects are included; a FE model using supplementary SPR indicators; a lagged FE model where SPR_{it} are replaced by SPR_{it-2}; two subgroup FE models, where observations are divided into one-child and multi-child families.[19] Here, the PAI is statistically significant at the 1% level across all models, showing that a 1% change yields an increase in the CADI of 13.85% for multi-child families, 20.61% for one-child families, and 17.64%–21.31% in general. Likewise, the key SPR indicator is consistently significant, demonstrating that a change in prioritizing education-relevant items from the lowest to the highest rank enhances children's academic development by 3.42% for one-child families, 6.57% for multi-child families, and 2.88%–5.13% in general.[20] Further, another three SPR indicators are found influential in general, that is, the healthcare-, donation- and clothing-relevant spending priority, despite being negligible in Column (3) or (5).[21]

[18] The test result suggests that the disturbance in our first-stage model is heteroskedastic. The levels of all exogenous regressors including fixed effects are involved.

[19] The same estimation using the CFA-based DV is reported in Table 3A.3. In addition, we replace the CADI with its components and report corresponding main results in Table 3A.4.

[20] Since the ranking scale is 1–10, changing from the lowest to highest priority rank is nine-fold, so we multiplied the reported coefficients by 9 to account for an overall difference.

[21] The effects of prioritizing financial support and social donation on children's academic development is perhaps a result of parental altruism that some parents are more willing to invest in their children's

Table 3.3: Spending Priorities, Parent Advantages and Child Academic Development

	(1) Baseline	(2) Main SPR	(3) New SPR	(4) Main Lag	(5) One-Child	(6) Multi-Child
PAI	0.2131***	0.1764***	0.2014***	0.1966***	0.2061***	0.1385***
	(0.0201)	(0.0249)	(0.0278)	(0.0336)	(0.0324)	(0.0378)
diet	0.0013	0.0024	0.0043	0.0021	-0.0001	0.0047**
	(0.0014)	(0.0015)	(0.0029)	(0.0021)	(0.0023)	(0.0022)
transport	0.0009	0.0024*	0.0042*	0.0018	0.0014	0.0036*
	(0.0012)	(0.0014)	(0.0021)	(0.0018)	(0.0020)	(0.0021)
utility	-0.0004	0.0006	0.0016	0.0025	0.0019	-0.0002
	(0.0011)	(0.0013)	(0.0019)	(0.0016)	(0.0019)	(0.0018)
necessities	-0.0024**	0.0003	0.0002	0.0019	0.0010	-0.0005
	(0.0011)	(0.0013)	(0.0019)	(0.0017)	(0.0019)	(0.0017)
healthcare	0.0007	0.0028**	0.0030*	0.0031*	0.0017	0.0041***
	(0.0010)	(0.0012)	(0.0018)	(0.0018)	(0.0019)	(0.0015)
education	0.0032***	0.0057***	0.0053***	0.0049**	0.0038*	0.0073***
	(0.0011)	(0.0012)	(0.0019)	(0.0019)	(0.0020)	(0.0017)
donation	0.0033**	0.0029*	0.0035	0.0034*	0.0008	0.0046**
	(0.0013)	(0.0015)	(0.0022)	(0.0020)	(0.0023)	(0.0020)
insurance	-0.0012	-0.0014	-0.0012	0.0029	-0.0003	-0.0031
	(0.0013)	(0.0014)	(0.0021)	(0.0020)	(0.0021)	(0.0021)
clothing	0.0015	0.0036**	0.0018	0.0055***	0.0027	0.0044**
	(0.0014)	(0.0015)	(0.0023)	(0.0021)	(0.0023)	(0.0019)
other	-0.0023*	0.0003	(omitted)	-0.0003	0.0005	0.0000
	(0.0012)	(0.0013)		(0.0020)	(0.0020)	(0.0018)
parent_age		0.0018***	0.0020***	0.0020***	0.0028***	0.0011
		(0.0006)	(0.0007)	(0.0007)	(0.0007)	(0.0008)
child gender		-0.0466***	-0.0495***	-0.0531***	-0.0463***	-0.0445***
		(0.0050)	(0.0058)	(0.0069)	(0.0063)	(0.0072)
sleep		0.0024	0.0026	0.0026	-0.0004	0.0026
		(0.0026)	(0.0031)	(0.0034)	(0.0036)	(0.0035)
child health		-0.0030***	-0.0032***	-0.0034***	-0.0062***	-0.0013
		(0.0010)	(0.0011)	(0.0013)	(0.0013)	(0.0014)
edu_savings		0.0242***	0.0232***	0.0248***	0.0183***	0.0334***
		(0.0050)	(0.0061)	(0.0069)	(0.0068)	(0.0079)
marital		-0.0313**	-0.0416***	-0.0446***	-0.0278*	-0.0363**
		(0.0121)	(0.0144)	(0.0166)	(0.0150)	(0.0179)
parental practices		0.1140***	0.0966***	0.0831***	0.0564***	0.1601***
		(0.0147)	(0.0178)	(0.0199)	(0.0211)	(0.0213)
residence		0.0026	-0.0009	0.0002	0.0147*	-0.0080
		(0.0055)	(0.0070)	(0.0076)	(0.0078)	(0.0071)
absence		0.0277**	0.0201	0.0100	0.0199	0.0274*
		(0.0125)	(0.0132)	(0.0157)	(0.0190)	(0.0156)
key_school		0.0285***	0.0224***	0.0196***	0.0283***	0.0272***
		(0.0049)	(0.0061)	(0.0066)	(0.0066)	(0.0072)
reaction (passive)		0.0398**	0.0303	0.0377	0.0679***	0.0234
		(0.0175)	(0.0208)	(0.0234)	(0.0227)	(0.0267)
reaction (positive)		0.0679***	0.0663***	0.0644***	0.0851***	0.0559***
		(0.0069)	(0.0089)	(0.0100)	(0.0114)	(0.0092)
Constant	0.6266***	0.4138***	0.4279***	0.4352***	0.4258***	0.4175***
	(0.0028)	(0.0368)	(0.0446)	(0.0489)	(0.0474)	(0.0482)
Time FE	N	Y	Y	Y	Y	Y
Province FE	N	Y	Y	Y	Y	Y
Birth FE	N	Y	Y	Y	Y	Y
Num. of child. FE	N	Y	Y	Y	N	Y
Obs	10299	7811	5190	4015	3685	4125

human capital formation than others (Das 2007).

Table 3.3: *Continued*

	Baseline	Main SPR	New SPR	Main Lag	One-Child	Multi-Child
R squared	0.0191	0.1085	0.1078	0.1091	0.1466	0.1020

Notes: The PAI and SPR variables are centred. In Column (4), all SPR variables are lagged by two periods because the CFPS was conducted biennially. Robust standard errors shown in parentheses are clustered at the province-birth cohort level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Source: Created by authors using CFPS data.

With regard to control variables, several findings are noteworthy. At the child level, girls' academic development is approximately 4.5%–5.3% better than boys', and, compared to students at ordinary schools, attending key schools is accompanied by an increase of 1.96%–2.85% in the CADI. At the parent or household level, having savings for educational purposes is found to prompt child outcomes by 1.83% for one-child families, 3.34% for multi-child families, and 2.32%–2.48% in general; children whose parents are divorced or widowed have a 2.78%–4.46% lower CADI than their counterparts; a 1% increase in parental practices yields 8.31%–11.4% higher CADI in general, while the impact for multi-child families is 184% higher than for one-child families; compared to negative reactions, a positive reaction to children's unsatisfactory test scores increases the CADI by 5.59% in multi-child families, 8.51% in one-child families, and 6.44%–6.79% in general. Additionally, parents' average age and children's health status are found to be influential in three other models but not in the multi-child model. In sum, having younger parents and worse health status are associated with the lower CADI.

We examined conversion rates, (Columns (1)–(4) of Table 3.4), using four conversion factors in order: a) children's gender, b) if either parent received higher education, c) if households lived in urban areas, and d) if children attended a key school. Moreover, in Column (5), we replace the PAI with differences in the PAI, i.e., $PAI_{it} - PAI_{it-2}$,[22]

[22] To ensure sufficient variations, we restricted respondents to those who had been surveyed at least three times.

to examine if parents' self-advancement over time influences their children. As seen in the table, two conversion factors, i.e., tertiary education and key school, are found to differentiate effects for four spending priorities.[23] Urban residence also plays a role in three aspects, while being born a boy rates as just better at converting insurance-relevant resources into the CADI. More specifically, categories of children who have one highly-educated parent at least, convert utility-, necessity-, education- and transfer-relevant goods and services more efficiently into their development. Noteworthy is that a difference in the CADI can be as large as 21.96% arising from prioritizing education-relevant items from the lowest to the highest rank when a highly-educated parent is involved. Similarly, attending a key school further helps children to convert necessity-, healthcare-, education- and transfer-relevant resources into their academic performance. In contrast, compared to their rural counterparts, children living in urban areas convert healthcare-, education- and transfer-relevant resources relatively inefficiently. Lastly and importantly, as shown in Column (5), a 1% increase in the PAI growth raises the CADI by 7.11%. This finding reveals that what parents achieve for themselves is also influential to their children's development, at least academically.

Table 3.4: Estimates of Conversion Factors With Interaction Effects

	(1) Gender	(2) Tertiary	(3) Residence	(4) Key School	(5) PAI Growth
PAI	0.1745***	0.1721***	0.1781***	0.1775***	
	(0.0248)	(0.0250)	(0.0248)	(0.0252)	
PAI_growth					0.0711**
					(0.0315)
child gender	-0.0461***	-0.0464***	-0.0466***	-0.0465***	-0.0507***
	(0.0051)	(0.0050)	(0.0050)	(0.0050)	(0.0083)
parent_highedu		0.0185			
		(0.0138)			
residence	0.0028	0.0019	0.0011	0.0030	0.0099
	(0.0055)	(0.0055)	(0.0055)	(0.0055)	(0.0082)
key_school	0.0286***	0.0277***	0.0282***	0.0297***	0.0141*
	(0.0049)	(0.0049)	(0.0049)	(0.0048)	(0.0074)
diet	0.0031	0.0021	0.0040**	0.0015	0.0024
	(0.0022)	(0.0015)	(0.0019)	(0.0018)	(0.0045)
transport	0.0014	0.0023	0.0040**	-0.0000	0.0035

[23] For using the tertiary dummy, they are utility-, necessities-, education- and transfer-relevant spending items, while for using the key school dummy, they are transport-, necessities-, healthcare- and transfer-relevant spending items.

Table 3.4: *Continued*

	Gender	Tertiary	Residence	Key School	PAI_Growth
	(0.0019)	(0.0014)	(0.0018)	(0.0017)	(0.0032)
utility	-0.0002	0.0004	0.0014	-0.0003	0.0003
	(0.0017)	(0.0013)	(0.0018)	(0.0015)	(0.0030)
necessities	-0.0001	-0.0002	0.0015	-0.0012	0.0017
	(0.0016)	(0.0013)	(0.0017)	(0.0015)	(0.0028)
healthcare	0.0023	0.0030**	0.0052***	0.0013	0.0040
	(0.0016)	(0.0012)	(0.0015)	(0.0013)	(0.0028)
education	0.0049***	0.0051***	0.0082***	0.0044***	0.0054*
	(0.0019)	(0.0012)	(0.0016)	(0.0015)	(0.0032)
donation	0.0008	0.0024	0.0049**	0.0007	0.0035
	(0.0022)	(0.0015)	(0.0019)	(0.0016)	(0.0028)
insurance	-0.0054***	-0.0016	-0.0006	-0.0017	0.0003
	(0.0020)	(0.0015)	(0.0019)	(0.0017)	(0.0029)
clothing	0.0022	0.0036**	0.0037*	0.0035**	0.0041
	(0.0020)	(0.0015)	(0.0019)	(0.0016)	(0.0033)
others	0.0019	0.0004	0.0015	-0.0011	0.0003
	(0.0018)	(0.0013)	(0.0018)	(0.0016)	(0.0037)
diet&Dummy	-0.0009	0.0141	-0.0041	0.0032	
	(0.0032)	(0.0096)	(0.0032)	(0.0033)	
transport&Dummy	0.0023	0.0109	-0.0042	0.0095***	
	(0.0026)	(0.0078)	(0.0028)	(0.0027)	
utility&Dummy	0.0016	0.0163*	-0.0020	0.0029	
	(0.0025)	(0.0083)	(0.0027)	(0.0028)	
necessities&Dummy	0.0012	0.0198***	-0.0031	0.0054**	
	(0.0025)	(0.0075)	(0.0027)	(0.0024)	
healthcare&Dummy	0.0014	0.0057	-0.0062**	0.0061**	
	(0.0025)	(0.0075)	(0.0025)	(0.0024)	
education&Dummy	0.0018	0.0244**	-0.0067**	0.0053*	
	(0.0027)	(0.0096)	(0.0027)	(0.0030)	
donation&Dummy	0.0041	0.0217**	-0.0053*	0.0081***	
	(0.0029)	(0.0096)	(0.0029)	(0.0031)	
insurance&Dummy	0.0080***	0.0132	-0.0029	0.0012	
	(0.0028)	(0.0083)	(0.0028)	(0.0030)	
clothing&Dummy	0.0029	0.0096	-0.0006	-0.0002	
	(0.0029)	(0.0086)	(0.0031)	(0.0029)	
other&Dummy	-0.0028	0.0001	-0.0030	0.0051	
	(0.0028)	(0.0079)	(0.0027)	(0.0031)	
Constant	0.4136***	0.4135***	0.4148***	0.4144***	0.3862***
	(0.0369)	(0.0366)	(0.0370)	(0.0368)	(0.0753)
Other Controls	Y	Y	Y	Y	Y
Time FE	Y	Y	Y	Y	Y
Province FE	Y	Y	Y	Y	Y
Birth FE	Y	Y	Y	Y	Y
Num. of child. FE	Y	Y	Y	Y	Y
Obs	7811	7811	7811	7811	2709
R squared	0.1103	0.1110	0.1098	0.1105	0.1147

Notes: PAI and SPR variables are centred. Due to the table length, some control variables included in the estimation are not presented here. Robust standard errors shown in parentheses are clustered at the province-birth cohort level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Source: Created by authors using CFPS data.

3.6 Robustness Checks

Estimating an empty model is often the first step in initiating multilevel analyses. Thus, we report the results of two empty models in Table 3.5. As seen in the table, the intra-class correlation coefficient (ICC) is 0.3326 in Column (1), indicating that observations within the same household are similar to each other.[24] In contrast, the household-level variance (σ_{0jk}) indicates considerable between-cluster heterogeneity. In Column (2), we added "provinces" as a new level and estimated a three-level empty model. The ICC becomes slightly larger in Column (2) than it is in Column (1), and the household-level variance is accordingly smaller.

In Columns (3)–(4), we included all variables and time fixed-effects and use lagged SPR indicators in Column (4). Here, two SPR predictors associated with education- and clothing-relevant items are statistically significant in both models. In particular, a change in the CADI between prioritizing spending on education, culture and recreation, and travel from the lowest to the highest rank could be 3.69% and 4.68%. The magnitude of the effects revealed here is very close to the FE estimates. Further, having savings for children's educational purposes produces an increase of 2.17% and 2.26% in the CADI. In addition, at the parent level, for a 1% increase in the PAI (parental practices), the CADI goes up by 12.06% (13.49%) and 13.56% (10.96%); a passive (positive) reaction to children's unsatisfactory test scores is found to improve academic development by 3. 42% (6.14%) and 3.96% (6.25%); compared to those whose parents remained married, children with divorced or widowed parents achieve a 2.57% and 4.17% lower CADI. Moreover, at the child level, girls' CADI is on average 4.62% and 5.2% higher than boys'; having one more hours of sleep on weekdays increases the CADI by 1.02% and 1.19%; being less healthy dampens children's academic development; attending a key school corresponds to a 2.54% and 1.48% higher CADI.

We report estimates using Lewbel's (2012) IV in Columns (5)–(6), the latter of which adopts supplementary SPR variables. Both results confirm that the education-relevant

[24] ICC is calculated as the ratio of the between-group variance relative to the total variance in the sample. It describes the extent to which observations within city groups are similar to each other.

spending priority is influential to children's academic development and the magnitudes of the effects uncovered here are very close to that of our main specification. Besides it, only the SPR predictor associated with transport-relevant items is consistently statistically significant. Further, the results of Hansen's J test for over-identification point to a 14.45% and 21.74% chance of a type one error if the null is rejected, confirming the joint validity of the instruments.

Overall, comparing nested and IV results to those in Table 3.3, we conclude that a) fewer SPR indicators are found to be statistically significant here, but the education-relevant SPR predictor exhibits good robustness with different estimators, and b) regarding other variables, despite nuances in the magnitude, the vast majority of results are quite similar.

Table 3.5: Random Intercept Models and Heteroscedasticity-based IV Models

	(1) Household	(2) Prov-House	(3) Three Level	(4) Lagged	(5) IV	(6) IV (New SPR)
Household Level						
diet			0.0008	0.0020	0.0024	0.0043
			(0.0016)	(0.0018)	(0.0016)	(0.0030)
transport			0.0019	0.0018	0.0024*	0.0042**
			(0.0018)	(0.0015)	(0.0014)	(0.0019)
utility			0.0003	0.0024	0.0006	0.0016
			(0.0018)	(0.0015)	(0.0013)	(0.0018)
necessities			-0.0005	0.0017	0.0003	0.0002
			(0.0014)	(0.0016)	(0.0013)	(0.0020)
healthcare			0.0017	0.0021	0.0028**	0.0031
			(0.0011)	(0.0016)	(0.0012)	(0.0019)
education			0.0041***	0.0052***	0.0057***	0.0054***
			(0.0013)	(0.0019)	(0.0013)	(0.0020)
donation			0.0027**	0.0027	0.0029*	0.0035
			(0.0010)	(0.0018)	(0.0015)	(0.0022)
insurance			-0.0018	0.0036***	-0.0014	-0.0012
			(0.0018)	(0.0014)	(0.0014)	(0.0021)
clothing			0.0031**	0.0051*	0.0036**	0.0019
			(0.0014)	(0.0028)	(0.0014)	(0.0022)
other			0.0001	0.0002	0.0003	(omitted)
			(0.0012)	(0.0023)	(0.0014)	
edu_savings			0.0217***	0.0226***	0.0242***	0.0232***
			(0.0065)	(0.0073)	(0.0049)	(0.0060)
residence			0.0067	0.0047	0.0026	-0.0008
			(0.0065)	(0.0092)	(0.0045)	(0.0056)
Parent Level						
PAI			0.1206***	0.1356***	0.1764***	0.2013***
			(0.0214)	(0.0260)	(0.0228)	(0.0274)
parent_age			0.0004	0.0007**	0.0018***	0.0020***
			(0.0004)	(0.0003)	(0.0005)	(0.0005)
marital			-0.0257**	-0.0417***	-0.0313***	-0.0417***
			(0.0120)	(0.0137)	(0.0114)	(0.0139)
parental practices			0.1349***	0.1096***	0.1143***	0.0968***
			(0.0132)	(0.0212)	(0.0152)	(0.0186)
reaction (passive)			0.0342***	0.0396**	0.0398**	0.0303
			(0.0132)	(0.0161)	(0.0174)	(0.0212)

Table 3.5: *Continued*

	Household	Prov-House	Three Level	Lagged	IV	IV (New SPR)
reaction (positive)			0.0614***	0.0625***	0.0679***	0.0663***
			(0.0059)	(0.0062)	(0.0068)	(0.0086)
Child Level						
child gender			-0.0462***	-0.0520***	-0.0466***	-0.0496***
			(0.0054)	(0.0068)	(0.0040)	(0.0050)
sleep			0.0102***	0.0119***	0.0024	0.0026
			(0.0025)	(0.0032)	(0.0023)	(0.0029)
child health			-0.0021***	-0.0029***	-0.0030***	-0.0032***
			(0.0008)	(0.0010)	(0.0010)	(0.0011)
absence			0.0147	0.0037	0.0277**	0.0201
			(0.0126)	(0.0141)	(0.0116)	(0.0138)
key_school			0.0254***	0.0148***	0.0285***	0.0224***
			(0.0057)	(0.0050)	(0.0047)	(0.0058)
Intercept	0.6241***	0.6289***	0.3854***	-1.9058***		
	(0.0021)	(0.0061)	(0.0289)	(0.0323)		
ICC	0.3326	0.3357	0.3058	0.3085		
	(0.0115)	(0.0179)	(0.0273)	(0.0412)		
Random Effects						
σ_{00j}		0.0007	0.0004	0.0003		
		(0.0003)	(0.0001)	(0.0002)		
σ_{0jk}	0.0117	0.0111	0.0092	0.0096		
	(0.0005)	(0.0007)	(0.0008)	(0.0013)		
σ_{ijk}	0.0235	0.0235	0.0217	0.0221		
	(0.0004)	(0.0009)	(0.0011)	(0.0014)		
Over-identification Test						
Hansen J-statistics					664.61	541.84
Chi^2 p-value					0.1445	0.2174
Time FE	N	N	Y	Y	Y	Y
Province FE	N	N	N	N	Y	Y
Birth FE	N	N	N	N	Y	Y
Num. of child. FE	N	N	N	N	Y	Y
Obs	12928	12920	7812	4072	7812	5190

Notes: The PAI and SPR variables are centred. In Column (4), all SPR variables are lagged by two periods because the CFPS was conducted biennially. σ_{00j} is the province-level variance. σ_{0jk} is the household-level variance. σ_{ijk} is the individual-level variance. The last SPR coefficient in Column (6) is automatically omitted due to collinearities. Robust standard errors are shown in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Source: Created by authors using CFPS data.

3.7 Conclusion

Children's human development is often evaluated as if the priorities, defined by those who care most about them, did not matter. By using the capability approach, we focus on the concept of ranking as a way of examining spending priorities and the effect of parental advantages on their children's development. Specifically, we assessed the role of parental spending priorities on their children's academic development using a sample of 8,422 Chinese children and adolescents surveyed during 2012 to 2018. We found that families with the highest spending priority on their children's education achieved returns in academic performance ranging from 2.88 to 6.57% higher than those who gave the lowest priority to educational spending. Similarly, prioritizing clothing and healthcare could yield maximum returns in child development ranging from 3.24–4.95% and 2.52–3.69%, respectively. When accounting for parent's higher education attainment, we found that the implied academic performance gains increased to 21.96%. Results were further validated in hierarchical linear models and 2SLS models with Lewbel's (2012) instruments.

Although the linkages between household educational expenditures and child development had been previously explored in the literature, this study brings new evidence to light, based on the capability approach. The core outcome in our view is how parents evaluate their children's academic abilities can be influenced by their spending priorities. We used the capability approach as a way of talking about "reasons to value", linked to parental practices and spending prioritization. This echoes Sen (2017) social choice work and the determination of the relevant informational spaces for evaluating people's advantages. The analysis employs a broader version of the capability approach, not restricted to capabilities, and encompassing plural informational spaces. It goes further to consider not simply capabilities *per se* but how people value those capabilities, and the concept of ranking used for counting these values shows how people order their priorities in this regard. The impact of the rankings is clearly seen in the results reported in this chapter.

Thus, our study provides novel insights into how parents can achieve comprehensive outcomes in their children's academic development. It goes beyond a consideration of parental practices and children's test scores. How parents convert their resources to enhance children's capabilities can be captured, not merely by how much they spend, but by the extent to which parents prioritize their children's education and culture over other expenditure considerations. Notably, other spending categories, such as medication, healthcare, sports, clothing and beauty, was also found to influence child development outcomes, to some degree. On top of all the discussed novelties, our results, on one hand, are still harmonized with a rich strand of literature that examines the effects of household investments and parental practices on child development (e.g., Heckman and Mosso 2014; Vasilyeva et al. 2018). On the other hand, our quantification approach for assessing the value households place on different consumption categories extends to the ongoing debate on the measurement of spending priorities (e.g., Costa Filho and Rocha 2022; Ratigan 2017; Rudra 2007). Overall, the findings suggest that policy efforts could be put into enhancing the value of education particularly among disadvantaged households and providing training on how to create real opportunities for children to exercise agency. Lastly, it is crucial to recognize the potential existence of unobservable heterogeneity in real opportunities that could influence the outcomes of priorities calculated here.

Concluding Remarks

Rooted in human development, this dissertation explores human mobility and child development in China by examining a) the role of job prospects in labor migration decisions, b) intra-city mobility trajectories of 368 cities and how interventions in high-risk areas influence different travel behaviors, and c) relationships between household spending priorities and the development of children reflected in their learning outcomes and learning processes.

By introducing reference dependence, derived from prospect theory, to the dynamic discrete choice modeling framework, **Chapter 1** creates two proxies for employment and wage prospects. It then constructs a unique quasi-panel built upon the 2017 CHFS where 10,254 migrants and 56,173 natives are eventually included and longitudinal statistics of 283 cities during 1997–2017. We find that migration probabilities raise by 1.281–2.185 percentage points with a 10% surge in the ratio of sector-based employment prospects in cities of destination to cities of origin, with stronger effects observed for larger ratios. In contrast, wage prospects are found to have little impact. Individuals with a family migration network are approximately 6 percentage points more likely to migrate. Additionally, labor migrants are more likely to be male, unmarried, younger, and more educated. For empirical analyses, a variety of econometric methods, including monadic and dyadic fixed-effects estimators, multilevel logistic regression, and GMM estimation, are adopted. Overall, it contributes to the debate on migration decisions concerned with future attractiveness, the RUM model of migration, job opportunities versus urban amenities, and Chinese labor migration.

As the first study to assess short-term, weekly variations through the Phillips and Sul's (2007, 2009) advanced clustering technique, **Chapter 2** combines the latest Baidu

Mobility Data and the risk-level data of 368 cities to compile a distinctive panel. The results suggest that the distance in intra-city mobility is decreasing within clusters but increasing between different clusters during the transitional period, with cities such as Lanzhou, Xining, Lhasa, and Ürümqi falling further behind. A distinct disparity is observed between Western China and the rest of the country, both in terms of intra-city mobility and home-workplaces (HW) commutes. Further, the "high risk" alert including the following containment measures consistently had a greater impact on commuting rates than on dining, leisure, and recreational (DLR) travel rates during the studied sample period. The estimation of causal effects is performed by Callaway and Sant'Anna's (2021) DiD. Apart from the novel application, it further adds to the broader body of literature on spatiotemporal mobility and (post-)pandemic socio-economic patterns.

Chapter 3 provides a composite analysis of children's academic development by developing a series of innovative indicators in the light of the Capability Approach and to reflect Amartya Sen's emphasis on informational pluralism. 8,422 children and adolescents aged between 6–16 are retrieved from the CFPS's 2012, 2014, 2016, and 2018 waves. Our findings demonstrate that enhancing parent advantages by 1% leads to a 13.85% to 21.31% improvement in children's academic development. The difference in prioritizing spending on education, from the lowest to the highest, explains the growth in child outcomes ranging from 2.88% to 6.57%. Furthermore, prioritizing spending on healthcare and clothing is also influential to educational development, with the largest gain of 2.52–3.69% and 3.24–4.95%. For robustness checks, hierarchical linear modeling and heteroskedasticity-based IVs are considered. Being the first application of the Capability Approach to the development of Chinese children, this work extends the frontiers of the Capability Approach and goes beyond the limited focus on material dimensions into how families value education and their children's future.

Discussions and Policy Implications

China's extensive, rapid development of infrastructure and transportation boosts the circulation of human capital and labor forces between regions. A high-speed rail now takes only five hours to complete 1318 kilometers between Shanghai and Beijing. These factors underlie the persistence of high labor mobility into the future. At the same time, migrants were found to be increasingly older and more educated from 2000 to 2015, according to China's Migrant Population Development Reports. Significantly, only 2% of migrants had at least a college degree in 1990, while 25 years later, the percentage is 23.3%. Skilled migrants tend to prioritize career prospects over the quality of life in the migration decision-making process (Liu and Shen 2014). This trend well signifies why we should re-examine economic incentives from the angle of job prospects. It is evident that China's current and future cohorts, who are more educated, are more likely to make visionary migration decisions than their elder generations. The likelihood of getting a job and how good it is are the most common questions with which (potential) workers are concerned. A variety of dimensions can contribute to assessing job quality, and what we have attempted to highlight in this chapter is but one.

Individual fulfillment and aspirations that are much more personal and sophisticated (Becker and Teney 2020) can also affect how people regard their job prospects across regions. The point made in **Chapter 1** is basic and general. Its findings, as a whole, suggest that for small- and medium-sized cities that are less attractive to labor migrants than large cities, concentrating resources to stimulate the acceleration of employment in certain sectors of interest could help, and downward migration may even happen in this case.[25] This can be seen as developing "comparative advantages" in playing the game. On one hand, by doing so, cities can outperform others in certain fields, attracting workforce and enabling spillovers. On the other hand, given China's economic slowdown, large cities tend to stabilize at a lower growth rate compared to the past, while small- and medium-sized cities have the potential to develop at a faster pace than large cities. As studied in Becker and Teney (2020), future research could adopt a mixed-method approach to delve into a deeper analysis, aiming to understand the more personal and subjective beliefs of labor migrants. This would enrich the evaluative dimension of job prospects. When data become available, investigating whether and how social network dynamics, including virtual communities (e.g., Komito 2011), shape individual perceptions of job prospects across different occupations, industries, and/or locations and accordingly their migration decisions is also a valuable direction. Furthermore, since the reference-dependent migration model developed in this thesis allows for repeated moves, it would be intriguing to utilize appropriate data to test the model.

Further, the role of population movement in mobilizing economic recovery (Spelta and Pagnottoni 2021) and accelerating the spread of COVID-19 (Chakraborty and Mukherjee 2023) has brought both advantages and challenges to cities with different characteristics as the Zero-COVID Policy draws to a close. Travel-intensive cities, on one hand, are often more economically developed and can be more resilient to economic

[25]More broadly speaking, as illustrated in prospect theory, policy-makers can devise programs in an incremental manner to attract or retain talents. For example, they could split the rewards for high-skilled immigrants into several installments, with each payment being higher than the previous one. Presently, some cities, like Shenzhen, already distribute subsidies over a span of multiple years, but the installments are almost all paid at a fixed rate. In such cases, recipients will not perceive gains once their reference points have been updated.

crises. On the other hand, the risk of contagion their residents face remains relatively high, which could result in worse long-term health outcomes, especially among the elderly, if the SARS-CoV-2 virus does not disappear completely in the near future. In contrast, people in Western China may be safer in this sense, but the implied worrisome prospect of economic activities in these regions also needs to be kept an eye on. To conclude, **Chapter 2** demonstrates that a) one-size-fits-all policies may not be ideal, and post-pandemic travel policy-making can be developed case-by-case at the cluster level and tailored to their specific needs, particularly for the lowest two clusters of which cities are mainly in Xinjiang and Tibet, and b) given that lifting or removing stringent containment measures is not effective enough, progressive, supportive policies are needed, such as discounts for local transportation and subsidies for the vulnerable population to promote mobility and consumption, such that elementary remedies can be introduced immediately after that to prevent persistent side effects in affected cities or areas while balancing health and economic performance. In addition, cities with bustling local trips but insufficient medical resources need greater attention to take care of their population health.

In sum, **Chapter 2** argues for a more nuanced approach to post-pandemic travel policy-making, with a focus on cluster-level considerations and the promotion of progressive measures to address the specific challenges faced by different regions. Future research could compare long-term health outcomes between high-risk and low-risk areas to see if COVID-19 exacerbates population health on a wider scope and examine whether long COVID-19 outcomes are correlated with human mobility. The question of whether and to what extent persistent declines in commuting rates in affected cities can be attributed to work from home also needs to be answered in the future as data become available. In this regard, conducting semi-structured interviews can provide valuable insights for researchers seeking to understand personal experiences in high-risk areas. This approach can help demystify any undocumented factors that contribute to the sustained decrease in commuting levels in affected cities, which go beyond the shift to remote work and individuals' reluctance to travel. Delventhal et al. (2022) discuss potential changes in urban areas resulting from permanent increases in working from home, such as longer commuting distances and lower real estate prices. Despite being unclear if changes in remote work are truly permanent or not, other pandemics are likely to occur and are predicted to happen in the future. Therefore, learning from this pandemic will help prepare the response and adaptation to future pandemics.

Last but not least, as China advances to the status of a developed nation, its economy will require increasingly-higher skilled labor and, thus, a better-educated society (Borsi et al. 2022). The state's long-term development vision and efforts toward boosting social mobility and promoting common prosperity are reflected in the banning of private after-school tutoring and the closure of independent colleges. The former was perceived to give an unfair advantage to wealthier families for whom private tutoring was more affordable, and independent colleges, prone to predatory practices, were considered as a lower-quality alternative to higher education. Nevertheless, our results suggest that equalizing educational expenditure opportunities may not be enough. Households that are more willing to invest in children's education cannot be simply regarded as wealthier. Rather, as the results here reflect, such households could be those that value education more. Considering Confucianism, which formed the core of traditional Chinese culture, places great emphasis on education and academic achievement (Gu 2006), Chinese parents who overlook or give up cultivating their children's education are likely to need help far beyond monetary assistance. This finding has significant implications for future educational policy. The reasons behind some households falling behind in prioritizing their children's education may be explained, in part, by a) traditional gender norms shaping beliefs that spending on a daughter's education is a waste, which, although changing as China modernizes, could still bring lower expectations for girls than for boys (e.g., Chi and Rao 2003; Liu 2006), b) parents holding a negative attitude towards the usefulness of study, possibly because they themselves are illiterate or less-well educated, and c) an exaggerated emphasis on children's agency, in that children are expected to achieve outcomes by themselves through their own diligence and intelligence without the need for parents to engage in developing real opportunities for their children to exercise this agency (Biggeri et al. 2011a).

Given the newly-unveiled Three-Child Policy, both parents' perceptions and the actions consequent to their perceptions for the development of children's capabilities are now even more important for China, especially in view of the finding that the difference in children's academic performance could be enhanced by up to 6.57% in multiple-child families. As manifested in **Chapter 3**, the value placed on children's development and future can be reflected in the order of relevant resources devoted to it. Therefore, policies that reinforce personal and household values and stimulate parents to create real opportunities for their children to exercise agency will provide additional returns in China's human capital accumulation. In the end, it is important to acknowledge the existence of unobservable heterogeneity in real opportunities that drive household prioritization. For example, families residing in remote rural areas with limited access to educational resources may naturally assign lower priority to education, despite their willingness to invest more in their children. Carrying out fieldwork to gather both qualitative and quantitative data, such as through a self-report study that explores the extent to which parents believe that they have done their best to cultivate their children, would be valuable for expanding and deepening our analyses.

References

Alderson P. (Jan. 2001). "Research by children". In: *International Journal of Social Research Methodology* 4.2, pp. 139–153.

An R., Tong Z., Liu X., Tan B., Xiong Q., Pang H., Liu Y., and Xu G. (2023). "Post COVID-19 pandemic recovery of intracity human mobility in Wuhan: Spatiotemporal characteristic and driving mechanism". In: *Travel Behaviour and Society* 31, pp. 37–48.

Anand S. and Sen A. (Feb. 2000). "The income component of the human development index". In: *Journal of Human Development* 1.1, pp. 83–106.

Anich R., Biggeri M., Libanora R., and Mariani S. (2011). "Street children in Kampala and NGOs' actions: Understanding capabilities deprivation and expansion". In: *Children and the capability approach*. London: Palgrave Macmillan, pp. 107–136.

Arunachalam S., Bahadir S. C., Bharadwaj S. G., and Guesalaga R. (Sept. 2020). "New product introductions for low-income consumers in emerging markets". In: *Journal of the Academy of Marketing Science* 48.5, pp. 914–940.

Bai C., Yan H., Yin S., Feng C., and Wei Q. (2021). "Exploring the development trend of internet finance in China: Perspective from club convergence". In: *The North American Journal of Economics and Finance* 58, p. 101505.

Baldwin R. E. (Feb. 2001). "Core-periphery model with forward-looking expectations". In: *Regional Science and Urban Economics* 31.1, pp. 21–49.

Ballet J., Biggeri M., and Comim F. (2011). "Children's agency and the capability approach: A conceptual framework". In: *Children and the capability approach*. London: Palgrave Macmillan, pp. 22–45.

Baum C. F. and Lewbel A. (2019). "Advice on using heteroskedasticity-based identification". In: *The Stata Journal* 19.4, pp. 757–767.

Baumann R., Svec J., and Sanzari F. (June 2015). "The Relationship between Net Migration and Unemployment: The Role of Expectations". In: *Eastern Economic Journal* 41.3, pp. 443–458.

Beck N. (Sept. 2018). "Estimating grouped data models with a binary dependent variable and fixed effects: What are the issues". In: *arXiv preprint arXiv:1809.06505*.

Becker G. S. (Nov. 1974). "A theory of social interactions". In: *Journal of Political Economy* 82.6, pp. 1063–1093.

Becker G. S. and Tomes N. (Aug. 1976). "Child endowments and the quantity and quality of children". In: *Journal of Political Economy* 84.4, Part 2, S143–S162.

Becker R. and Teney C. (July 2020). "Understanding high-skilled intra-European migration patterns: the case of European physicians in Germany". In: *Journal of Ethnic and Migration Studies* 46.9, pp. 1737–1755.

Beine M., Bertoli S., and Fernández-Huertas Moraga J. (Apr. 2016). "A Practitioners' Guide to Gravity Models of International Migration". In: *The World Economy* 39.4, pp. 496–512.

Beine M., Bourgeon P., and Bricongne J.-C. (Jan. 2019). "Aggregate Fluctuations and International Migration". In: *The Scandinavian Journal of Economics* 121.1, pp. 117–152.

Beine M. and Parsons C. (Apr. 2015). "Climatic Factors as Determinants of International Migration". In: *The Scandinavian Journal of Economics* 117.2, pp. 723–767.

Bellanca N., Biggeri M., and Marchetta F. (July 2011). “An extension of the capability approach: Towards a theory of dis-capability”. In: *Alter* 5.3, pp. 158–176.

Bertoli S., Brücker H., and Fernández-Huertas Moraga J. (Sept. 2016). “The European crisis and migration to Germany”. In: *Regional Science and Urban Economics* 60, pp. 61–72.

Bertoli S. and Fernández-Huertas Moraga J. (May 2013). “Multilateral resistance to migration”. In: *Journal of Development Economics* 102, pp. 79–100.

Biggeri M., Ballet J., and Comim F. (2011a). *Children and the Capability Approach*. New York: Palgrave Macmillan.

— (2011b). “The capability approach and research on children: Capability approach and children’s issues”. In: *Children and the Good Life*. Springer, Dordrecht, pp. 75–89.

Biggeri M. and Mehrotra S. (2011). “Child poverty as capability deprivation: How to choose domains of child well-being and poverty”. In: *Children and the Capability Approach*. London: Palgrave Macmillan UK, pp. 46–75.

Binder M. and Broekel T. (May 2011). “Applying a non-parametric efficiency analysis to measure conversion efficiency in Great Britain”. In: *Journal of Human Development and Capabilities* 12.2, pp. 257–281.

— (Aug. 2012a). “Happiness no matter the cost? An examination on how efficiently individuals reach their happiness levels”. In: *Journal of Happiness Studies* 13.4, pp. 621–645.

— (Jan. 2012b). “The neglected dimension of well-being: Analyzing the development of “conversion efficiency” in Great Britain”. In: *The Journal of Socio-Economics* 41.1, pp. 37–47.

Black M. M., Trude A. C., and Lutter C. K. (Sept. 2020). “All children thrive: Integration of nutrition and early childhood development”. In: *Annual Review of Nutrition* 40.1, pp. 375–406.

Blundell R. and Bond S. (Nov. 1998). “Initial conditions and moment restrictions in dynamic panel data models”. In: *Journal of Econometrics* 87.1, pp. 115–143.

Borghans L., Meijers H., and Weel B. ter (Jan. 2008). “The role of noncognitive skills in explaining cognitive test scores”. In: *Economic Inquiry* 46.1, pp. 2–12.

Borsi M. T., Valerio Mendoza O. M., and Comim F. (Feb. 2022). “Measuring the provincial supply of higher education institutions in China”. In: *China Economic Review* 71, p. 101724.

Burchardt T. and Vizard P. (Feb. 2011). “‘Operationalizing’ the capability approach as a basis for equality and human rights monitoring in twenty-first-century Britain”. In: *Journal of Human Development and Capabilities* 12.1, pp. 91–119.

Burton P., Phipps S., and Curtis L. (Apr. 2002). “All in the family: A simultaneous model of parenting style and child conduct”. In: *American Economic Review* 92.2, pp. 368–372.

Byskov M. F. (2018). “Selecting cpabilities for development”. In: *New frontiers of the capability approach*. Ed. by F. Comim, S. Fennell, and P. B. Anand. Cambridge University Press, pp. 198–231.

Callaway B. and Sant’Anna P. H. (2021). “Difference-in-differences with multiple time periods”. In: *Journal of Econometrics* 225.2, pp. 200–230.

Caselli F., Grigoli F., and Sandri D. (2021). “Protecting lives and livelihoods with early and tight lockdowns”. In: *The BE Journal of macroeconomics* 22.1, pp. 241–268.

Chakraborty T. and Mukherjee A. (2023). “Economic geography of contagion: a study of COVID-19 outbreak in India”. In: *Journal of Population Economics*, pp. 1–33.

Chen A. and Groenewold N. (2019). “China’s ‘New Normal’: Is the growth slowdown demand- or supply-driven?” In: *China Economic Review* 58, p. 101203.

Chen C. and Fan C. C. (2016). “China’s Hukou Puzzle: Why Don’t Rural Migrants Want Urban Hukou?” In: *China Review* 16.3, pp. 9–39.

Chen R., Zhang M., and Zhou J. (2023). “Jobs-housing relationships before and amid COVID-19: An excess-commuting approach”. In: *Journal of Transport Geography* 106, p. 103507.

Chen S. H., Deng X. F., Zhang E., Wang L. K., and Liu C. H. (Nov. 2021). "Self-regulatory development in children from Chinese immigrant families: Evidence for commonality and specificity". In: *Child Development* 92.6.

Cheng Z., Nielsen I., and Smyth R. (2014). "Access to social insurance in urban China: A comparative study of rural–urban and urban–urban migrants in Beijing". In: *Habitat International* 41, pp. 243–252.

Chi J. and Rao N. (Sept. 2003). "Parental beliefs about school learning and children's educational attainment: Evidence from rural China". In: *Ethos* 31.3, pp. 330–356.

Chung S. H., Zhang M., and Partridge M. D. (2020). "Positive feedback in skill aggregation across Chinese cities". In: *Regional Studies* 54.11, pp. 1596–1611.

Cicala S. (2022). "JUE: Insights: Powering work from home". In: *Journal of Urban Economics*, p. 103474.

Clark W. A. V. and Lisowski W. (Sept. 2017). "Prospect theory and the decision to move or stay". In: *Proceedings of the National Academy of Sciences* 114.36, E7432–E7440.

Colin Cameron A. and Miller D. L. (Apr. 2015). "A Practitioner's Guide to Cluster-Robust Inference". In: *Journal of Human Resources* 50.2, pp. 317–372.

Comim F. (Dec. 2021). "A poset-generalizability method for human development indicators". In: *Social Indicators Research* 158.3, pp. 1179–1198.

Comim F., Fennell S., and Anand P. B., eds. (Oct. 2018). *New Frontiers of the Capability Approach*. Cambridge University Press.

Cooper K. and Stewart K. (June 2021). "Does household income affect children's outcomes? A systematic review of the evidence". In: *Child Indicators Research* 14.3, pp. 981–1005.

Costa Filho M. C. and Rocha A. C. (2022). "Meanings and attitudes regarding education and household spending priorities of the new middle-class families in Brazil". In: *Journal of Consumer Culture* 22.1, pp. 183–206.

Cui Y., Liu H., and Zhao L. (Sept. 2019). "Mother's education and child development: Evidence from the compulsory school reform in China". In: *Journal of Comparative Economics* 47.3, pp. 669–692.

Cunha F. and Heckman J. J. (Apr. 2007). "The technology of skill formation". In: *American Economic Review* 97.2, pp. 31–47.

Czaika M. (2015). "Migration and Economic Prospects". In: *Journal of Ethnic and Migration Studies* 41.1, pp. 58–82.

Czaika M. and Parsons C. R. (Apr. 2017). "The Gravity of High-Skilled Migration Policies". In: *Demography* 54.2, pp. 603–630.

Dai R., Mookherjee D., Munshi K., Zhang X., et al. (2019). *The community origins of private enterprise in China*. Tech. rep. Boston University-Department of Economics.

Das M. (Sept. 2007). "Persistent inequality: An explanation based on limited parental altruism". In: *Journal of Development Economics* 84.1, pp. 251–270.

De Chaisemartin C. and d'Haultfoeuille X. (2020). "Two-way fixed effects estimators with heterogeneous treatment effects". In: *American Economic Review* 110.9, pp. 2964–2996.

De Haas H. (2009). "Mobility and Human Development". In: *UNDP (United Nations Development Programme)*.

De Jong G. F. (Jan. 2000). "Expectations, gender, and norms in migration decision-making". In: *Population Studies* 54.3, pp. 307–319.

Delventhal M., Kwon E., and Parkhomenko A. (2022). "JUE Insight: How do cities change when we work from home?" In: *Journal of Urban Economics* 127. JUE Insights: COVID-19 and Cities, p. 103331.

Dooley M. and Stewart J. (Feb. 2007). “Family income, parenting styles and child behavioural-emotional outcomes”. In: *Health Economics* 16.2, pp. 145–162.

Fang G., Tang T., Zhao F., and Zhu Y. (2023). “The social scar of the pandemic: Impacts of COVID-19 exposure on interpersonal trust”. In: *Journal of Asian Economics* 86, p. 101609.

Fang H., Wang L., and Yang Y. (Nov. 2020). “Human mobility restrictions and the spread of the Novel Coronavirus (2019-nCoV) in China”. In: *Journal of Public Economics* 191, p. 104272.

Foltz J., Guo Y., and Yao Y. (June 2020). “Lineage networks, urban migration and income inequality: Evidence from rural China”. In: *Journal of Comparative Economics* 48.2, pp. 465–482.

Frey C. B. and Osborne M. A. (2017). “The future of employment: How susceptible are jobs to computerisation?” In: *Technological Forecasting and Social Change* 114, pp. 254–280.

Fu Y. and Gabriel S. A. (2012). “Labor migration, human capital agglomeration and regional development in China”. In: *Regional Science and Urban Economics* 42.3, pp. 473–484.

Ge P., Sun W., and Zhao Z. (May 2021). “Employment structure in China from 1990 to 2015”. In: *Journal of Economic Behavior & Organization* 185, pp. 168–190.

Gibbs H., Liu Y., Pearson C. A., Jarvis C. I., Grundy C., Quilty B. J., Diamond C., and Eggo R. M. (2020). “Changing travel patterns in China during the early stages of the COVID-19 pandemic”. In: *Nature communications* 11.1, p. 5012.

Glaeser E. L., Gorback C., and Redding S. J. (Jan. 2022). “JUE Insight: How much does COVID-19 increase with mobility? Evidence from New York and four other U.S. cities”. In: *Journal of Urban Economics* 127, p. 103292.

Goldhagen J. L., Shenoda S., Oberg C., Mercer R., Kadir A., Raman S., Waterston T., and Spencer N. J. (Jan. 2020). “Rights, justice, and equity: A global agenda for child health and wellbeing”. In: *The Lancet Child & Adolescent Health* 4.1, pp. 80–90.

Gong D., Shang Z., Su Y., Yan A., and Zhang Q. (2023). “A systematic evaluation of the economic impacts of China’s zero-COVID policies”. en. In: *SSRN Electron. J.*

Gu M. (June 2006). “An analysis of the impact of traditional Chinese culture on Chinese education”. In: *Frontiers of Education in China* 1.2, pp. 169–190.

Gunnar R. J., Kanerva K., Salmi S., Häyrinen T., Haataja L., Pakarinen M. P., and Merras-Salmio L. (Jan. 2020). “Neonatal intestinal failure is independently associated with impaired cognitive development later in childhood”. In: *Journal of Pediatric Gastroenterology & Nutrition* 70.1, pp. 64–71.

Hansen L. P. (July 1982). “Large Sample Properties of Generalized Method of Moments Estimators”. In: *Econometrica* 50.4, p. 1029.

Hanushek E. A. and Woessmann L. (Aug. 2008). “The role of cognitive skills in economic development”. In: *Journal of Economic Literature* 46.3, pp. 607–668.

Hanushek E. A., Rivkin S. G., and Schiman J. C. (Dec. 2016). “Dynamic effects of teacher turnover on the quality of instruction”. In: *Economics of Education Review* 55, pp. 132–148.

Hatzigeorgiou A. and Lodefalk M. (2021). “A literature review of the nexus between migration and internationalization”. In: *The Journal of International Trade & Economic Development* 30.3, pp. 319–340.

He C. and Gober P. (Dec. 2003). “Gendering Interprovincial Migration in China”. In: *International Migration Review* 37.4, pp. 1220–1251.

Heckman J. J. (July 2008a). “Role of income and family influence on child outcomes”. In: *Annals of the New York Academy of Sciences* 1136.1, pp. 307–323.

— (July 2008b). “Schools, skills, and synapses”. In: *Economic Inquiry* 46.3, pp. 289–324.

Heckman J. J. and Mosso S. (Aug. 2014). “The economics of human development and social mobility”. In: *Annual Review of Economics* 6.1, pp. 689–733.

Heckman J. J., Stixrud J., and Urzua S. (July 2006). "The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior". In: *Journal of Labor Economics* 24.3, pp. 411–482.

Horrace W. C. and Oaxaca R. L. (Mar. 2006). "Results on the bias and inconsistency of ordinary least squares for the linear probability model". In: *Economics Letters* 90.3, pp. 321–327.

Hsu S. and Wu Y.-Y. (2015). *Education as Cultivation in Chinese Culture*. Singapore: Springer.

Hu S., Xiong C., Liu Z., and Zhang L. (2021). "Examining spatiotemporal changing patterns of bike-sharing usage during COVID-19 pandemic". In: *Journal of transport geography* 91, p. 102997.

Hu S., Xiong C., Younes H., Yang M., Darzi A., and Jin Z. C. (2022). "Examining spatiotemporal evolution of racial/ethnic disparities in human mobility and COVID-19 health outcomes: Evidence from the contiguous United States". In: *Sustainable Cities and Society* 76, p. 103506.

Huang L., Xie R., and Yang G. (2022). "The impact of lockdown on air pollution: Evidence from an instrument". In: *China Economic Review* 71, p. 101731.

Iacus S. M., Santamaria C., Sermi F., Spyratos S., Tarchi D., and Vespe M. (2020). "Human mobility and COVID-19 initial dynamics". In: *Nonlinear Dynamics* 101.3, pp. 1901–1919.

Jiao J., Lee H. K., and Choi S. J. (2022). "Impacts of COVID-19 on bike-sharing usages in Seoul, South Korea". In: *Cities* 130, p. 103849.

Kahneman D. and Tversky A. (1979). "Prospect theory: An analysis of decision under risk. Econometrika, 47, 263-291." In: *Econometrica* 47.2, pp. 263–292.

Kennan J. and Walker J. R. (2011). "The Effect of Expected Income on Individual Migration Decisions". In: *Econometrica* 79.1, pp. 211–251.

Komito L. (2011). "Social media and migration: Virtual community 2.0". In: *Journal of the American society for information science and technology* 62.6, pp. 1075–1086.

Kőszegi B. and Rabin M. (May 2009). "Reference-Dependent Consumption Plans". In: *American Economic Review* 99.3, pp. 909–936.

Kraemer M. U., Yang C.-H., Gutierrez B., Wu C.-H., Klein B., Pigott D. M., Group† O. C.-1. D. W., Du Plessis L., Faria N. R., Li R., et al. (2020). "The effect of human mobility and control measures on the COVID-19 epidemic in China". In: *Science* 368.6490, pp. 493–497.

Krauss J. E., Artur L., Brockington D., Castro Jr E., Fernando Jr J., Fisher J., Kingman A., Moises H. M., Mlambo A., Nuvunga M., et al. (2022). "'To prevent this disease, we have to stay at home, but if we stay at home, we die of hunger'–Livelihoods, vulnerability and coping with Covid-19 in rural Mozambique". In: *World Development* 151, p. 105757.

Krumbiegel K., Maertens M., and Wollni M. (Feb. 2018). "The role of fairtrade certification for wages and job satisfaction of plantation workers". In: *World Development* 102, pp. 195–212.

Lansdown G. (2004). "Participation and young children". In: *Early Childhood Matters* 103, pp. 4–14.

Lee Y. J. and Xiao Z. (1998). "Children's support for elderly parents in urban and rural China: Results from a national survey". In: *Journal of Cross-Cultural Gerontology* 13.1, pp. 39–62.

Lei W., Alves L. G., and Amaral L. A. N. (2022). "Forecasting the evolution of fast-changing transportation networks using machine learning". In: *Nature communications* 13.1, p. 4252.

Lewbel A. (2012). "Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models". In: *Journal of Business & Economic Statistics* 30.1, pp. 67–80.

Li B. and Ma L. (2022). "JUE insight: Migration, transportation infrastructure, and the spatial transmission of COVID-19 in China". In: *Journal of urban economics* 127, p. 103351.

Li L. and Wu X. (Sept. 2019). "Housing price and intergenerational co-residence in urban China". In: *Journal of Housing Economics* 45, p. 101596.

Li T., Wang J., Huang J., Yang W., and Chen Z. (2021). "Exploring the dynamic impacts of COVID-19 on intercity travel in China". In: *Journal of Transport Geography* 95, p. 103153.

Li X. and Luo L. (Oct. 2021). "Migration patterns and migrant workers' consumption". In: *China Agricultural Economic Review* 13.4, pp. 781–798.

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

Liang Z. (Sept. 2001). "The Age of Migration in China". In: *Population and Development Review* 27.3, pp. 499–524.

Liebenehm S. (Feb. 2018). "Temporal stability of risk attitudes and the impact of adverse shocks—a panel data analysis from Thailand and Vietnam". In: *World Development* 102, pp. 262–274.

Liu D., Chen D., and Brown B. B. (Jan. 2020a). "Do parenting practices and child disclosure predict parental knowledge? A meta-analysis". In: *Journal of Youth and Adolescence* 49.1, pp. 1–16.

Liu F. (Sept. 2006). "Boys as only-children and girls as only-children—parental gendered expectations of the only-child in the nuclear Chinese family in present-day China". In: *Gender and Education* 18.5, pp. 491–505.

Liu S., Kong G., and Kong D. (2020b). "Effects of the COVID-19 on air quality: Human mobility, spillover effects, and city connections". In: *Environmental and Resource Economics* 76, pp. 635–653.

Liu Y., Li Z., and Breitung W. (2012). "The social networks of new-generation migrants in China's urbanized villages: A case study of Guangzhou". In: *Habitat International* 36.1, pp. 192–200.

Liu Y. and Shen J. (Nov. 2014). "Spatial patterns and determinants of skilled internal migration in China, 2000-2005". In: *Papers in Regional Science* 93.4, pp. 749–771.

Liu Y., Wang Z., Rader B., Li B., Wu C.-H., Whittington J. D., Zheng P., Stenseth N. C., Bjornstad O. N., Brownstein J. S., et al. (2021). "Associations between changes in population mobility in response to the COVID-19 pandemic and socioeconomic factors at the city level in China and country level worldwide: a retrospective, observational study". In: *The Lancet Digital Health* 3.6, e349–e359.

Liu Z. and Yu L. (2020). "Stay or Leave? The Role of Air Pollution in Urban Migration Choices". In: *Ecological Economics* 177.June 2019.

Long J. A. and Ren C. (2022). "Associations between mobility and socio-economic indicators vary across the timeline of the Covid-19 pandemic". In: *Computers, environment and urban systems* 91, p. 101710.

Lynch A. D., Lerner R. M., and Leventhal T. (Jan. 2013). "Adolescent academic achievement and school engagement: An examination of the role of school-wide peer culture". In: *Journal of Youth and Adolescence* 42.1, pp. 6–19.

Mayda A. M. (Sept. 2010). "International migration: a panel data analysis of the determinants of bilateral flows". In: *Journal of Population Economics* 23.4, pp. 1249–1274.

McFadden D. (1974). "Conditional Logit Analysis of Qualitative Choice Behaviour". In: *Frontiers in Econometrics*. Ed. by P. Zarembka. New York: Academic Press New York, pp. 105–142.

Meng L. (2020). "Permanent migration desire of Chinese rural residents: Evidence from field surveys, 2006–2015". In: *China Economic Review* 61.January, pp. 1–15.

Meng X. and Xue S. (Jan. 2020). "Social networks and mental health outcomes: Chinese rural–urban migrant experience". In: *Journal of Population Economics* 33.1, pp. 155–195.

Mitze T. and Kosfeld R. (2022). "The propagation effect of commuting to work in the spatial transmission of COVID-19". In: *Journal of geographical systems* 24.1, pp. 5–31.

Mu X., Yeh A. G.-O., Zhang X., Wang J., and Lin J. (May 2021). “Moving down the urban hierarchy: Turning point of China’s internal migration caused by age structure and hukou system”. In: *Urban Studies*, p. 004209802110077.

Mu X., Zhang X., Yeh A. G.-O., Yu Y., and Wang J. (2023). “Structural Changes in Human Mobility Under the Zero-COVID Strategy in China”. In: *Environment and Planning B: Urban Analytics and City Science*, p. 23998083231159397.

Munshi K. (2020). “Social networks and migration”. In: *Annual Review of Economics* 12, pp. 503–524.

Nussbaum M. (Mar. 2000). *Women and Human Development*. Cambridge: Cambridge University Press.

— (2011). *Creating Capabilities: The Human Development Approach*. Cambridge: Harvard University Press.

Nussbaum M. C. (2018). *The monarchy of fear: A philosopher looks at our political crisis*. New York: Simon & Schuster.

Nussbaum M. C. (2006). *Frontiers of justice: Disability, nationality, species membership*. Cambridge: Belknap Press.

Ortega F. and Peri G. (Mar. 2013). “The effect of income and immigration policies on international migration”. In: *Migration Studies* 1.1, pp. 47–74.

Pagan A. R. and Hall A. D. (1983). “Diagnostic tests as residual analysis”. In: *Econometric Reviews* 2.2, pp. 159–218.

Peng L. and Swider S. (Jan. 2017). “Migration and regional inequality: changing characteristics of China’s economic inequality”. In: *Eurasian Geography and Economics* 58.1, pp. 89–113.

Peng Y. (2004). “Kinship networks and entrepreneurs in China’s transitional economy”. In: *American Journal of Sociology* 109.5, pp. 1045–1074.

Pesaran M. H. (July 2006). “Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure”. In: *Econometrica* 74.4, pp. 967–1012.

Phillips P. C. and Sul D. (2007). “Transition modeling and econometric convergence tests”. In: *Econometrica* 75.6, pp. 1771–1855.

— (2009). “Economic transition and growth”. In: *Journal of applied econometrics* 24.7, pp. 1153–1185.

Poncet S. (May 2006). “Provincial migration dynamics in China: Borders, costs and economic motivations”. In: *Regional Science and Urban Economics* 36.3, pp. 385–398.

Punch S. (2002). “Negotiating autonomy: childhoods in rural Bolivia”. In: *Conceptualising child-adult relations*. Ed. by L. Alanen and B. Mayall. London: Routledge, pp. 37–50.

Qu Y., Cai F., and Zhang X. (2013). “Has the ‘Flying Geese’ Occured in China? An Analysis of China’s Manufacturing Industries from 1998 to 2009”. In: *China Economic Quarterly* 12.3, pp. 757–776.

Ratigan K. (2017). “Disaggregating the developing welfare state: Provincial social policy regimes in China”. In: *World Development* 98, pp. 467–484.

Rawls J. (Dec. 1971). *A Theory of Justice*. Cambridge: Harvard University Press.

Ren G., Zhu X., Heerink N., and Feng S. (2020). “Rural household migration in China – the roles of actual and perceived tenure security”. In: *China Economic Review* 63.6, p. 101534.

Roodman D. (2009). “How to do xtabond2: An introduction to difference and system GMM in Stata”. In: *Stata Journal* 9.1, pp. 86–136.

Rose L. T. and Fischer K. W. (2011). “Intelligence in childhood”. In: *The Cambridge handbook of intelligence*. Cambridge: Cambridge University Press, pp. 144–174.

Rosenzweig M. and Zhang J. (Apr. 2014). *Co-residence, life-cycle savings and inter-generational support in urban China*. Tech. rep. Cambridge, MA: National Bureau of Economic Research.

Royuela V. and Ordóñez J. (June 2018). “Internal migration in a developing country: A panel data analysis of Ecuador (1982-2010)”. In: *Papers in Regional Science* 97.2, pp. 345–367.

Rudra N. (2007). “Welfare states in developing countries: Unique or universal?” In: *the Journal of Politics* 69.2, pp. 378–396.

Şahin A., Tansel A., and Berument M. H. (July 2015). “Output-Employment Relationship across Sectors: A Long- versus Short-Run Perspective”. In: *Bulletin of Economic Research* 67.3, pp. 265–288.

Saito M. (Feb. 2003). “Amartya Sen’s capability approach to education: A critical exploration”. In: *Journal of Philosophy of Education* 37.1, pp. 17–33.

Schmidt-Catran A. W. and Spies D. C. (Apr. 2016). “Immigration and Welfare Support in Germany”. In: *American Sociological Review* 81.2, pp. 242–261.

Sen A. (1980). “Equality of what?” In: *The Tanner Lectures on Human Values*. Vol. 1. Cambridge: Cambridge University Press, pp. 197–220.

— (July 1997). “Maximization and the act of choice”. In: *Econometrica* 65.4, p. 745.

— (1999). “Development as Freedom”. In: Oxford: Oxford University Press.

— (2002). *Rationality and Freedom*. Harvard University Press.

— (2009). *The Idea of Justice*. Cambridge: Harvard University Press.

— (2017). *Collective Choice and Social Welfare*. London: Penguin Books.

Shrestha M. (Feb. 2020). “Get Rich or Die Tryin’: Perceived Earnings, Perceived Mortality Rates, and Migration Decisions of Potential Work Migrants from Nepal”. In: *The World Bank Economic Review* 34.1, pp. 1–27.

Shrestha M., Weissman A., Thapa B., Adhikari R., and Perry K. (June 2021). “Malnutrition matters: Impact of nutritional status on early childhood development in Nepal”. In: *Current Developments in Nutrition* 5.Supplement_2, pp. 688–688.

Silverstein M. and Zhang W. (Apr. 2020). “Grandparents’ financial contributions to grandchildren in rural China: The role of remittances, household structure, and patrilineal culture”. In: *The Journals of Gerontology: Series B* 75.5. Ed. by D. Carr, pp. 1042–1052.

Small K. A. and Rosen H. S. (Jan. 1981). “Applied Welfare Economics with Discrete Choice Models”. In: *Econometrica* 49.1, p. 105.

Smith M. D., Rabbitt M. P., and Coleman- Jensen A. (May 2017). “Who are the world’s food insecure? New evidence from the Food and Agriculture Organization’s food insecurity experience scale”. In: *World Development* 93, pp. 402–412.

Song Y. (June 2014). “What should economists know about the current Chinese hukou system?” In: *China Economic Review* 29, pp. 200–212.

Spelta A. and Pagnottoni P. (2021). “Mobility-based real-time economic monitoring amid the COVID-19 pandemic”. In: *Scientific reports* 11.1, pp. 1–15.

Su Y., Tesfazion P., and Zhao Z. (Feb. 2018). “Where are the migrants from? Inter- vs. intra-provincial rural-urban migration in China”. In: *China Economic Review* 47, pp. 142–155.

Sun L. and Abraham S. (2021). “Estimating dynamic treatment effects in event studies with heterogeneous treatment effects”. In: *Journal of Econometrics* 225.2, pp. 175–199.

Sun R. (Sept. 2004). “Worry about medical care, family support, and depression of the elders in urban China”. In: *Research on Aging* 26.5, pp. 559–585.

Sunio V., Ugay J. C., Li C.-W., Liwanag H. J., and Santos J. (2023). “Impact of Public Transport Disruption on Access to Healthcare Facility and Well-being During the Covid-19 Pandemic: A Qualitative Case Study in Metro Manila, Philippines”. In: *Case Studies on Transport Policy*, p. 100948.

Suzumura K. (2011). In: *Handbook of Social Choice and Welfare*. Ed. by K. J. Arrow, A. K. Sen, and K. Suzumura. 1st ed. Vol. 2. Elsevier. Chap. 23, pp. 605–685.

Swift A. and Brighouse H. (2014). “Family values: The ethics of parent-child relationships”. In: Princeton: Princeton University Press.

Sylvia S., Luo R., Zhong J., Dill S.-E., Medina A., and Rozelle S. (Jan. 2022). “Passive versus active service delivery: Comparing the effects of two parenting interventions on early cognitive development in rural China”. In: *World Development* 149, p. 105686.

Tan Y., Xu H., and Zhang X. (June 2016). “Sustainable urbanization in China: A comprehensive literature review”. In: *Cities* 55, pp. 82–93.

Tian X., Zhang X., Zhou Y., and Yu X. (2016). “Regional income inequality in China revisited: A perspective from club convergence”. In: *Economic Modelling* 56.C, pp. 50–58.

Tokey A. I. (2021). “Spatial association of mobility and COVID-19 infection rate in the USA: A county-level study using mobile phone location data”. In: *Journal of Transport & Health* 22, p. 101135.

Tukey J. W. (1977). *Exploratory Data Analysis*. Reading, MA: Addison-Wesley.

United Nations Development Programme (2020). *Human Mobility, Shared Opportunities: A Review of the 2009 Human Development Report and the Way Ahead*. UNDP.

Valerio Mendoza O. M., Borsi M. T., and Comim F. (Apr. 2022). “Human capital dynamics in China: Evidence from a club convergence approach”. In: *Journal of Asian Economics* 79.C.

Vasilyeva M., Laski E., Veraksa A., Weber L., and Bukhalenkova D. (Aug. 2018). “Distinct pathways from parental beliefs and practices to children’s numeric skills”. In: *Journal of Cognition and Development* 19.4, pp. 345–366.

Wang L., Xue Y., Chang M., and Xie C. (2020). “Macroeconomic determinants of high-tech migration in China: The case of Yangtze River Delta Urban Agglomeration”. In: *Cities* 107.September 2019, p. 102888.

Windmeijer F. (May 2005). “A finite sample correction for the variance of linear efficient two-step GMM estimators”. In: *Journal of Econometrics* 126.1, pp. 25–51.

Winters P., De Janvry A., and Sadoulet E. (2001). “Family and community networks in Mexico-US migration”. In: *Journal of human resources*, pp. 159–184.

Wintoki M. B., Linck J. S., and Netter J. M. (Sept. 2012). “Endogeneity and the dynamics of internal corporate governance”. In: *Journal of Financial Economics* 105.3, pp. 581–606.

World Bank Group (2018). *World Development Report 2018: Learning to Realize Education’s Promise*. Washington, DC: World Bank.

Wu J., Zhan X., Xu H., and Ma C. (2023). “The economic impacts of COVID-19 and city lockdown: Early evidence from China”. In: *Structural Change and Economic Dynamics*.

Xie Y., Zhang X., Tu P., Ren Q., Sun Y., Lv P., Ding H., Hu J., and Wu Q. (2017). *China Family Panel Studies User’s Manual (3rd edition)*. Tech. rep. Institute of Social Science Survey, Peking University.

Xin R., Ai T., Ding L., Zhu R., and Meng L. (2022). “Impact of the COVID-19 pandemic on urban human mobility-A multiscale geospatial network analysis using New York bike-sharing data”. In: *Cities* 126, p. 103677.

Ye J., Wu X., and Tan J. (2016). “Migrate to skilled cities: Human capital agglomeration and urban-to-urban migration in China”. In: *Emerging Markets Finance and Trade* 52.8, pp. 1762–1774.

Yergeau M.-E. (Mar. 2020). “Tourism and local welfare: A multilevel analysis in Nepal’s protected areas”. In: *World Development* 127, p. 104744.

Yi D., Deng X., Fan G.-Z., and Ong S. E. (Oct. 2018). “House price and co-residence with older parents: Evidence from China”. In: *The Journal of Real Estate Finance and Economics* 57.3, pp. 502–533.

Yin T. (Dec. 2010). “Parent–child co-residence and bequest motives in China”. In: *China Economic Review* 21.4, pp. 521–531.

Young M. E. et al. (2002). “From early child development to human development”. In: *Washington, DC: World Bank*.

Yousefzadeh S., Biggeri M., Arciprete C., and Haisma H. (Apr. 2019). “A capability approach to child growth”. In: *Child Indicators Research* 12.2, pp. 711–731.

Zhang J., Huang J., Wang J., and Guo L. (2020). “Return migration and Hukou registration constraints in Chinese cities”. In: *China Economic Review* 63, p. 101498.

Zhang J., Wang R., and Lu C. (Mar. 2019). “A quantitative analysis of Hukou reform in Chinese cities: 2000-2016”. In: *Growth and Change* 50.1, pp. 201–221.

Zhang J. and Zhao Z. (2013). “Measuring the income-distance tradeoff for rural-urban migrants in China”. In: *IZA Discussion Paper* 7160.

Zhang Z., Ma C., and Wang A. (July 2021). “A longitudinal study of multidimensional poverty in rural China from 2010 to 2018”. In: *Economics Letters* 204, p. 109912.

Zhao X. (2020). “Migrants and urban wage: Evidence from China’s internal migration”. In: *China Economic Review* 61.March, p. 101287.

Zhou G., Chu G., Li L., and Meng L. (Jan. 2020). “The effect of artificial intelligence on China’s labor market”. In: *China Economic Journal* 13.1, pp. 24–41.

Zhu H. (Jan. 2016). “Adult children’s characteristics and intergenerational financial transfers in urban China”. In: *Chinese Journal of Sociology* 2.1, pp. 75–94.

Zhu J. and Lin B. (2020). “Convergence analysis of city-level energy intensity in China”. In: *Energy Policy* 139, p. 111357.

Appendix 1

Appendix 1A: Regression Tables of Complete Results

Table 1A.1: Determinants of Migration Decisions (1997–2017): Employment Prospects with Fixed Effects

	OLS		Multilateral Resistance to Migration						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
distance_ETrend	0.0011***	0.0020***	0.0014***	0.0018***	0.0015***	0.0017***	0.0013***	0.0013***	0.0018***
	(0.0003)	(0.0003)	(0.0003)	(0.0007)	(0.0003)	(0.0005)	(0.0003)	(0.0003)	(0.0005)
gender		0.0009***	0.0006***	0.0005***	0.0002	0.0006***	0.0006***	0.0006***	0.0006***
		(0.0001)	(0.0001)	(0.0002)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
marriage		-0.0080***	-0.0074***	-0.0048***	-0.0041***	-0.0072***	-0.0074***	-0.0074***	-0.0073***
		(0.0006)	(0.0005)	(0.0006)	(0.0005)	(0.0005)	(0.0005)	(0.0005)	(0.0005)
hukou_type		-0.0010**	-0.0004	-0.0008**	0.0000	-0.0004	-0.0004	-0.0004	-0.0004
		(0.0004)	(0.0003)	(0.0003)	(0.0002)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
health_status		-0.0002**	-0.0001	0.0000	-0.0001*	-0.0001*	-0.0001	-0.0001*	-0.0001*
		(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
hh_income		0.0000	-0.0000	0.0001	-0.0001**	0.0000	-0.0000	-0.0000	0.0000
		(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
age		-0.0025***	-0.0023***	-0.0016***	-0.0014***	-0.0023***	-0.0023***	-0.0023***	-0.0023***
		(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
age^2		0.0000***	0.0000***	0.0000***	0.0000***	0.0000***	0.0000***	0.0000***	0.0000***
		(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
schooling		0.0002***	0.0002***	0.0002***	0.0000	0.0002***	0.0002***	0.0002***	0.0002***
		(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
pioneer		0.0667***	0.0619***	0.0575***	0.0666***	0.0596***	0.0614***	0.0617***	0.0601***
		(0.0043)	(0.0035)	(0.0063)	(0.0045)	(0.0033)	(0.0036)	(0.0037)	(0.0034)
distance_lngdppc		0.0144***	0.0003	0.0034	0.0038	-0.0019	-0.0016	-0.0020	-0.0027
		(0.0023)	(0.0044)	(0.0062)	(0.0045)	(0.0055)	(0.0045)	(0.0045)	(0.0054)
distance_coop		-0.0004	-0.0007**	-0.0003	-0.0007**	-0.0008**	-0.0008***	-0.0008***	-0.0008*
		(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0004)	(0.0003)	(0.0003)	(0.0004)
distance_medical		2.2892**	3.2870***	2.4882**	3.0634***	4.1929***	3.2512***	3.2223***	4.1993***
		(1.0015)	(0.9410)	(0.9847)	(0.8152)	(1.2921)	(1.0355)	(1.0390)	(1.2887)
distance_highEdu		0.1441*	0.0961	-0.0415	0.0404	0.1681*	0.1215	0.1262	0.1716*
		(0.0812)	(0.0922)	(0.1496)	(0.0871)	(0.0941)	(0.0948)	(0.0903)	(0.0886)
distance_ppDen		0.0026	-0.0016	-0.0008	0.0008	-0.0097	-0.0007	-0.0006	-0.0096
		(0.0019)	(0.0084)	(0.0055)	(0.0074)	(0.0075)	(0.0080)	(0.0080)	(0.0074)
distance_tertiaryRatio		0.0024	-0.0348	-0.0259	-0.0380	-0.0306	-0.0335	-0.0368	-0.0320
		(0.0188)	(0.0254)	(0.0342)	(0.0249)	(0.0273)	(0.0253)	(0.0252)	(0.0271)
distance_CHRI				0.0221**					
				(0.0088)					
distance_ETrendXdistance_CHRI				0.0022**					
				(0.0010)					
Constant	0.0089***	0.0622***	0.0604***	0.0405***	0.0396***	0.0607***	0.0600***	0.0601***	0.0606***
	(0.0009)	(0.0045)	(0.0041)	(0.0038)	(0.0042)	(0.0042)	(0.0042)	(0.0042)	(0.0041)
Time FE	Y	Y	Y	Y	Y				
Industry FE			Y	Y	Y	Y	Y		
Origin FE			Y	Y		Y			Y
Destination FE			Y	Y			Y	Y	
Pairs of cities FE					Y				
Origin-year FE							Y	Y	
Dest-year FE						Y			Y
Industry-year FE								Y	Y
R^2	0.0014	0.0459	0.0534	0.0509	0.0842	0.0721	0.0697	0.0696	0.0721
Obs	969998	749219	729965	404487	729960	729769	729725	748981	749034

Notes: Trending indicators estimated here are untransformed, i.e., $E\_Trending_{ik,t} - E\_Trending_{ij,t}$. Standard errors shown in parentheses are clustered at the destination city. Coefficients are displayed as 0.0000 because the values are smaller than 0.0001. The small R^2 is endemic to and a direct result of our discrete choice setting. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Source: Created by authors using CHFS, CHRI, and China Data Institute (2021).

Table 1A.2: Determinants of Migration Decisions: Multilevel Logit and Two-step System GMM

	Two Level Logit			Three Level Logit			GMM		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
distance_ETrend	0.0533***	0.0446***	0.0325***	0.0534***	0.0445***	0.0325***	0.3157***	0.2493***	0.2751***
	(0.0106)	(0.0058)	(0.0104)	(0.0106)	(0.0058)	(0.0104)	(0.0754)	(0.0569)	(0.0582)
gender	0.1153***	0.0808***	0.0303*	0.1265***	0.0950***	0.0303*	0.0010***	0.0010***	0.0010***
	(0.0238)	(0.0192)	(0.0159)	(0.0243)	(0.0194)	(0.0159)	(0.0002)	(0.0002)	(0.0002)
marriage	-0.4273***	-0.3765***	-0.4047***	-0.4110***	-0.3520***	-0.4047***	-0.0078***	-0.0076***	-0.0076***
	(0.0302)	(0.0346)	(0.0242)	(0.0300)	(0.0340)	(0.0242)	(0.0006)	(0.0006)	(0.0006)
hukou_type	-0.0421	-0.0344	0.0332	-0.0589*	-0.0567	0.0332	-0.0008*	-0.0007*	-0.0008*
	(0.0350)	(0.0442)	(0.0240)	(0.0355)	(0.0447)	(0.0240)	(0.0004)	(0.0004)	(0.0004)
health_status	-0.0163	-0.0083	0.0004	-0.0165	-0.0056	0.0004	-0.0002**	-0.0002*	-0.0002*
	(0.0147)	(0.0135)	(0.0112)	(0.0148)	(0.0136)	(0.0112)	(0.0001)	(0.0001)	(0.0001)
hh_income	0.0012	-0.0066	-0.0148**	-0.0020	-0.0076	-0.0148**	0.0001	0.0001	0.0001
	(0.0099)	(0.0080)	(0.0075)	(0.0099)	(0.0083)	(0.0075)	(0.0001)	(0.0001)	(0.0001)
age	-0.1944***	-0.1885***	-0.1943***	-0.1930***	-0.1889***	-0.1943***	-0.0024***	-0.0023***	-0.0023***
	(0.0077)	(0.0080)	(0.0065)	(0.0074)	(0.0083)	(0.0065)	(0.0002)	(0.0002)	(0.0002)
age^2	0.0019***	0.0018***	0.0022***	0.0018***	0.0018***	0.0022***	0.0000***	0.0000***	0.0000***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0000)	(0.0000)	(0.0000)
schooling	0.0366***	0.0311***	0.0052	0.0406***	0.0461***	0.0052	0.0003***	0.0002***	0.0002***
	(0.0051)	(0.0061)	(0.0039)	(0.0052)	(0.0070)	(0.0039)	(0.0000)	(0.0000)	(0.0000)
pioneer	1.7622***	1.5286***	1.1741***	1.8058***	1.5418***	1.1741***	0.0713***	0.0705***	0.0692***
	(0.0657)	(0.1024)	(0.0415)	(0.0670)	(0.1093)	(0.0415)	(0.0042)	(0.0041)	(0.0042)
distance_lngdppc	0.8749***	0.5826***	0.2135***	0.8861***	0.5819***	0.2135***	0.0133***	0.0086**	0.0107**
	(0.0888)	(0.0934)	(0.0376)	(0.0881)	(0.0948)	(0.0376)	(0.0032)	(0.0041)	(0.0045)
distance_ coop	-0.0129	0.0068	-0.0073*	-0.0121	0.0069	-0.0073*	-0.0005**	-0.0004	-0.0005
	(0.0082)	(0.0144)	(0.0039)	(0.0082)	(0.0143)	(0.0039)	(0.0003)	(0.0003)	(0.0003)
distance_medical	46.4214*	43.7924	26.8754**	45.4386	45.3497	26.8756**	2.5699***	2.2949**	2.6314**
	(27.9812)	(42.9833)	(12.4029)	(27.9493)	(43.6284)	(12.4029)	(0.8988)	(0.9026)	(1.0154)
distance_highEdu	3.8617	6.5296	0.4295	3.9861	6.7124	0.4295	0.1378	0.2271**	0.2377**
	(2.4409)	(4.4566)	(0.7884)	(2.4537)	(4.4735)	(0.7884)	(0.0893)	(0.0882)	(0.0919)
distance_ppDen	0.0464	0.1925**	-0.0033	0.0475	0.1953**	-0.0033	0.0046	0.0159***	0.0160***
	(0.0771)	(0.0809)	(0.0205)	(0.0773)	(0.0819)	(0.0205)	(0.0043)	(0.0050)	(0.0049)
distance_tertiaryRatio	0.2462	1.8194**	-0.1408	0.2182	1.7686**	-0.1408	-0.0269	-0.0673**	-0.0847**
	(0.5814)	(0.7838)	(0.2204)	(0.5774)	(0.7913)	(0.2204)	(0.0300)	(0.0323)	(0.0347)
Constant	-1.8388***	-1.5829***	0.0083	-1.9502***	-1.7668***	0.0083	0.0612***	0.0593***	0.0593***
	(0.3101)	(0.2573)	(0.2468)	(0.3137)	(0.2857)	(0.2468)	(0.0046)	(0.0046)	(0.0045)
Level 2 variance	0.1484	1.6617	1.2686	0.1870	0.2416	0.0000			
	(0.0372)	(0.1672)	(0.1010)	(0.0312)	(0.0455)	(0.0000)			
Level 3 variance			0.1054	1.6046	1.2686				
			(0.0398)	(0.1685)	(0.1010)				
ICC	0.0432	0.3356	0.2783	0.0816	0.3595	0.2783			
	(0.0104)	(0.0224)	(0.0160)	(0.0112)	(0.0214)	(0.0160)			
Time FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Industry FE							Y	Y	Y
Num. of instruments							182	180	201
AR(2)							0.772	0.831	0.824
Hansen's J test							0.368	0.278	0.408
Difference-in-Hansen tests									
GMM instruments for levels – Excluding group							0.123	0.103	0.113
GMM instruments for levels – Difference (null H = exogenous)							0.466	0.377	0.622
GMM instrument for distance_trend – Excluding group							0.553	0.381	0.521
GMM instrument for distance_trend – Difference (null H = exogenous)							0.103	0.179	0.185
Obs	749219	749219	749219	749219	749219	749219	729965	729965	729965

Notes: Trending indicators estimated here are untransformed, i.e., $E\_Trending_{ik,t} - E\_Trending_{ij,t}$. Standard errors shown in parentheses are clustered at the destination city, except for multilevel logit models, where robust standard errors are applied but not clustered. The Windmeijer correction (Windmeijer 2005) is enabled in the GMM estimation. Coefficients are displayed as 0.0000 because the values are smaller than 0.0001. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Source: Created by authors using CHFS and China Data Institute (2021).

Appendix 1B: Regression Tables of Complementary Results

Table 1B.1: Determinants of Migration Decisions (1997–2017): Employment Growth Rates

	OLS		Multilateral Resistance to Migration			
	(1)	(2)	(3)	(4)	(5)	(6)
distance_EGrowth	0.0035***	0.0006	0.0002	0.0006	0.0005	0.0001
	(0.0008)	(0.0005)	(0.0004)	(0.0003)	(0.0006)	(0.0005)
gender		0.0009***	0.0006***	0.0001	0.0006***	0.0006***
		(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
marriage		-0.0079***	-0.0074***	-0.0041***	-0.0072***	-0.0074***
		(0.0006)	(0.0005)	(0.0005)	(0.0005)	(0.0005)
hukou_type		-0.0010**	-0.0004*	-0.0000	-0.0004*	-0.0004*
		(0.0004)	(0.0003)	(0.0002)	(0.0003)	(0.0003)
health_status		-0.0002**	-0.0001*	-0.0001*	-0.0001*	-0.0001
		(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
hh_income		0.0000	-0.0000	-0.0001**	0.0000	-0.0000
		(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
age		-0.0024***	-0.0023***	-0.0014***	-0.0023***	-0.0023***
		(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
age^2		0.0000***	0.0000***	0.0000***	0.0000***	0.0000***
		(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
schooling		0.0002***	0.0002***	0.0000	0.0002***	0.0002***
		(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
pioneer		0.0670***	0.0622***	0.0667***	0.0600***	0.0617***
		(0.0044)	(0.0036)	(0.0046)	(0.0034)	(0.0036)
distance_lngdppc		0.0141***	-0.0011	0.0023	-0.0017	-0.0031
		(0.0023)	(0.0045)	(0.0048)	(0.0055)	(0.0045)
distance_coop		-0.0004	-0.0007**	-0.0007**	-0.0008**	-0.0009***
		(0.0003)	(0.0003)	(0.0003)	(0.0004)	(0.0003)
distance_medical		2.2484**	3.2036***	2.9437***	4.0403***	3.2219***
		(0.9989)	(0.9130)	(0.7792)	(1.2303)	(1.0067)
distance_highEdu		0.1465*	0.1033	0.0485	0.1771*	0.1322
		(0.0815)	(0.0940)	(0.0885)	(0.0964)	(0.0957)
distance_ppDen		0.0025	-0.0020	0.0005	-0.0099	-0.0008
		(0.0019)	(0.0083)	(0.0073)	(0.0075)	(0.0079)
distance_tertiaryRatio		0.0034	-0.0374	-0.0410	-0.0291	-0.0365
		(0.0188)	(0.0264)	(0.0261)	(0.0272)	(0.0264)
Constant	0.0088***	0.0616***	0.0600***	0.0394***	0.0601***	0.0597***
	(0.0009)	(0.0045)	(0.0042)	(0.0042)	(0.0042)	(0.0042)
Time FE	Y	Y	Y	Y		
Industry FE			Y	Y	Y	Y
Origin FE			Y		Y	
Destination FE			Y			Y
Pairs of cities FE				Y		
Origin-year FE						Y
Dest-year FE					Y	
R^2	0.0016	0.0458	0.0534	0.0840	0.0722	0.0698
Obs	1038407	758754	739308	739303	739109	739067

Notes: The variable "distance_JobGrowth" is defined as $GR_{ik,t} - GR_{ij,t}$. Standard errors shown in parentheses are clustered at the destination city. Coefficients are displayed as 0.0000 because the values are smaller than 0.0001. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Source: Created by authors using CHFS and China Data Institute (2021).

Table 1B.2: Determinants of Migration Decisions (1997–2017): Wage Prospects with Fixed Effects

	OLS		Multilateral Resistance to Migration					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
distance_wage	0.0237***	0.0041***	0.0039***	0.0033**	0.0002	0.0004	-0.0000	0.0002
	(0.0024)	(0.0014)	(0.0013)	(0.0014)	(0.0013)	(0.0013)	(0.0015)	(0.0013)
gender		0.0003***	0.0006***	0.0000	0.0002**	-0.0001	0.0003**	0.0002**
		(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
marriage		-0.0041***	-0.0043***	-0.0039***	-0.0040***	-0.0023***	-0.0040***	-0.0040***
		(0.0005)	(0.0005)	(0.0005)	(0.0004)	(0.0003)	(0.0004)	(0.0004)
hukou_type		-0.0005	-0.0005**	-0.0002	-0.0001	0.0001	-0.0002	-0.0001
		(0.0003)	(0.0002)	(0.0003)	(0.0002)	(0.0001)	(0.0002)	(0.0002)
health_status		-0.0002*	-0.0001	-0.0001	-0.0001	-0.0001	-0.0001	-0.0001
		(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
hh_income		0.0002*	0.0002***	0.0001	0.0001	-0.0001	0.0001*	0.0001*
		(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
age		-0.0021***	-0.0020***	-0.0020***	-0.0019***	-0.0010***	-0.0019***	-0.0019***
		(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0001)	(0.0001)	(0.0001)
age^2		0.0000***	0.0000***	0.0000***	0.0000***	0.0000***	0.0000***	0.0000***
		(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
schooling		0.0002***	0.0002***	0.0002***	0.0002***	0.0000	0.0002***	0.0002***
		(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
pioneer		0.0785***	0.0794***	0.0737***	0.0744***	0.0778***	0.0735***	0.0743***
		(0.0046)	(0.0044)	(0.0041)	(0.0040)	(0.0052)	(0.0039)	(0.0041)
distance_unemploy		-0.0845	-0.0766	-0.0808	-0.1569	-0.1479	-0.1943*	-0.2129
		(0.0970)	(0.0926)	(0.0979)	(0.1124)	(0.1154)	(0.1054)	(0.1310)
distance_coop		0.0002	0.0001	0.0002	-0.0006	-0.0006	-0.0005	-0.0006*
		(0.0003)	(0.0002)	(0.0003)	(0.0004)	(0.0004)	(0.0005)	(0.0003)
distance_medical		3.1689***	3.3611***	3.0711***	1.6373**	1.2017	2.2222**	1.4692**
		(0.8900)	(0.8360)	(1.0386)	(0.6755)	(0.7307)	(0.9063)	(0.7346)
distance_highEdu		0.2206***	0.1963***	0.2220***	0.0491	-0.0043	0.0803	0.1016*
		(0.0689)	(0.0632)	(0.0751)	(0.0632)	(0.0626)	(0.0780)	(0.0614)
distance_ppDen		0.0042**	0.0036*	0.0056***	-0.0037	-0.0041	-0.0114**	-0.0028
		(0.0018)	(0.0018)	(0.0019)	(0.0080)	(0.0082)	(0.0051)	(0.0078)
distance_tertiaryRatio		0.0070	-0.0046	0.0218	-0.0162	-0.0256	-0.0027	-0.0135
		(0.0169)	(0.0157)	(0.0173)	(0.0201)	(0.0217)	(0.0205)	(0.0219)
Constant	0.0074***	0.0478***	0.0470***	0.0457***	0.0460***	0.0293***	0.0458***	0.0456***
	(0.0007)	(0.0034)	(0.0034)	(0.0033)	(0.0030)	(0.0031)	(0.0030)	(0.0031)
Time FE	Y	Y	Y	Y	Y	Y	Y	Y
Industry FE			Y	Y	Y	Y	Y	Y
Origin FE			Y		Y	Y	Y	
Destination FE				Y	Y			Y
Pairs of cities FE						Y		
Origin-year FE								Y
Dest-year FE							Y	
R^2	0.0072	0.0418	0.0442	0.0460	0.0501	0.0851	0.0671	0.0683
Obs	1050567	664285	664284	664283	664282	664212	664121	664074

Notes: Standard errors shown in parentheses are clustered at the destination city. Coefficients are displayed as 0.0000 because the values are smaller than 0.0001. The small R^2 is endemic to and a direct result of our discrete choice setting. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Source: Created by authors using CHFS, CHRI, and China Data Institute (2021).

Appendix 1C: Additional Figures

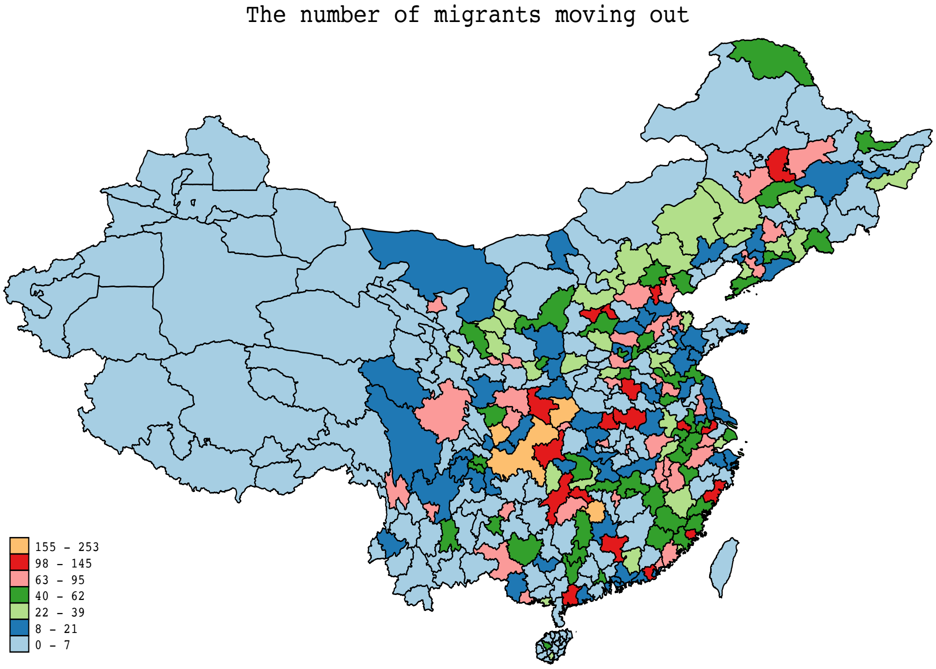


Figure 1C.1: Geographic distribution of emigrants across Chinese prefecture-level cities.

Source: Author's elaboration using CHFS.

Appendix 1D: Data Wrangling Report

Introduction

The China Household Financial Survey was conducted biannually between 2011 and 2017. The latest wave tracked the majority of the preceding households, as well as newly-surveyed families, for a total of 127,012 individuals involved.[1] Each wave has three sets of data that provide information on individuals, households, and cities. We used the latest wave because information on individuals' origin cities is absent in earlier waves. This latest wave satisfies all the requirements of our approach: (a) it includes both natives and migrants and (b) allows us to compile origin information at the prefecture city level.

It should be noted that, as a household survey, respondents were asked to provide information about their family members. The respondent acted as a delegate of his or her family, and the CHFS's interviewers chose the person who was most familiar with his or her household's economic conditions to answer the questions. However, this might not always be feasible. In other words, the respondent could either be him- or herself or share certain family relationships with the persons who were indicated (parent, spouse, etc.). This design entails a few concerns. For example, when the respondents were answering on their own behalf, their resident cities are not clearly stated in the data since the question only asked where their family members were living now, if they did not live together. Instead, we can learn in which city respondents got surveyed from the city table. However, where people were surveyed is not always the city they mostly lived

[1]Some households were surveyed previously but lost in the later waves.

in. We identified this discrepancy in two ways. Firstly, the 2017 questionnaire asked if the current city/county was the place where the family's main economic activities were carried out. An individual who did not stay with family members but visited or went back when the survey was conducted can be a respondent, although intuitively they should not be. Secondly, two variables helped us check if individuals lived in their Hukou registration city/county/town. Based on them, we made a comparison between the Hukou registration city and surveyed city. A few thousand observations were found to be dubious. For instance, the variable A2019b shows that the individual lived in his or her Hukou registration city but he or she was surveyed in another city. To clean and compile the data, when surveyed cities were used as resident cities, we only retained householder samples.

Migrant Identification

```sql
case
   when a2019e is not null then 1
   when a2023g = 1 then 1
   when
         transfer = 0 and ((a2001 != 1 and a2016b_short != a2019_short) or
                     (a2001 = 1 and hhead = 1 and a2016b_short != a2019_short)) then
                     1
   when transfer = 1 and (((a2019_short != a2022m_short and a2019b is not null) or
                    (a2019b is null and
                     ((a2016b is null and surveyedCity != a2022m_short and
                      (a2001 != 1 or (a2001 = 1 and hhead = 1)))
                       or (a2016b is not null and a2016b_short != a2022m_short and
                         (a2001 != 1 or (a2001 = 1 and hhead = 1))))))) then 1
   else 0 end migrant,
```

Figure 1D.1: SQL codes for identifying migrants.

Overall, there are four types of migrants. First, based on the variable A2019e, we distinguished the floating population from natives and migrants with Hukou transfers.[2] The 2017 questionnaire specifies that the referring question would only be delivered if individuals' Hukou registration cities are different from their resident cities. As shown in Figure D1, we specified A2019e to be not null. However, sometimes the data can be messy and display information dissimilar to what is asked in the questionnaire, as shown in Figure D2. To avoid mistreating certain migrants as natives resulting from missing inputs in A2019e, we coded a supplementary condition starting with transfer=0.[3] Another type that we can straightforwardly identify is returnees. When the CHFS system detected that the individual's Hukou registration city is identical to his/her resident city, the interviewer then moved on to question A2023g – if he/she had ever left the Hukou registration city for somewhere else for more than six months. It is worth noting that although interviewers only asked family members aged above 16 in 2017, returnees can be quite young when they migrated. In our final dataset, we dropped these observations.

Identifying two other types is more challenging. We can determine whether an individual was living in their Hukou registration city or not by examining variable

[2]This question asked "in which year did he/she leave the city where his/her Hukou is registered". All the variables discussed in this report are summarized in Table 1D.1.

[3]Variables marked with "short" indicate their inputs are based on the first four digits of the NBS codes. This is to ensure that migrants are identified at the prefecture level.

290 rows

	hhid	pline	a2001	hhead	a2016b	a2019	a2019e
1	201103874	1	7	1	31-3101-310101	34-3412-341226	<null>
2	2015021551	3	6	<null>	52-5201-520102	41-4113-411323	<null>
3	2017000201	6	6	<null>	33-3300-330000	42-4210-421023	<null>
4	2017000415	2	2	<null>	31-3101-310107	32-3208-320803	<null>
5	2013013344	6	6	<null>	35-3506-350602	35-3507-350702	<null>
6	2013013344	5	6	<null>	35-3501-350102	35-3507-350702	<null>
7	2015035622	3	6	<null>	43-4300-430000	62-6201-620104	<null>
8	2015018175	5	2	1	12-1200-120000	35-3509-350923	<null>
9	2013020837	5	6	<null>	44-4420-442000	45-4504-450406	<null>
10	2013020837	3	6	<null>	44-4420-442000	45-4504-450406	<null>
11	201107732	4	6	<null>	11-1100-110000	51-5116-511603	<null>
12	2015028005	1	10	<null>	44-4419-441900	44-4452-445281	<null>
13	2015032364	5	8	<null>	32-3205-320500	51-5115-511521	<null>
14	2017001638	5	7	<null>	32-3202-320281	32-3212-321203	<null>
15	2017001853	6	6	<null>	44-4401-440100	46-4604-460400	<null>
16	201103848	1	6	1	31-3101-310113	34-3412-341226	<null>
17	201105086	12	7777	<null>	53-5325-532522	41-4113-411302	<null>

Figure 1D.2: Respondents whose Hukou registration cities and resident cities are mismatching.

Notes: These failed to be identified by A2019e.

	hhid	pline	a2001	a2016b_short	a2019_short	a2019b	transfer	a2022m_short	surveyedCity
1	2015014195	2	2	<null>	3101	<null>	1	3302	3302
2	2017000273	2	2	<null>	3415	<null>	1	5107	3101
3	2017000301	2	2	<null>	4414	<null>	1	1408	4401
4	2017000355	2	2	<null>	5105	<null>	1	5002	5001
5	2015013590	6	7	<null>	3311	<null>	1	3301	3301
6	2017000356	2	2	<null>	4415	<null>	1	4416	4403
7	2017000490	2	10	<null>	6104	<null>	1	6110	6101
8	201108103	5	6	<null>	6101	<null>	1	6103	6103
9	2017000512	1	1	<null>	3102	<null>	1	6529	3101
10	2017000530	2	2	<null>	1307	<null>	1	1310	1306
11	2017001086	5	6	<null>	4403	<null>	1	4409	4409
12	2015035430	2	2	<null>	6211	<null>	1	6201	6201
13	2013019521	4	6	<null>	4401	<null>	1	4408	4408
14	2017001293	4	6	<null>	4406	<null>	1	4401	4401
15	2017001429	2	2	<null>	3503	<null>	1	3505	3505
16	2015025396	3	2	<null>	4406	<null>	1	4401	4401
17	2017000842	3	7	<null>	5107	<null>	1	<null>	5101
18	2017001293	2	2	<null>	4406	<null>	1	4401	4401
19	2017001441	1	1	<null>	4403	<null>	1	2201	2102

Figure 1D.3: Migrants with Hukou transfer: Surveyed cities are identical to origin cities.

A2019e, but there were some migrants with Hukou transfers that require additional variables to locate. Specifically, we used A2022k (referred to as "transfer" in our codes) and A2022m to identify four cases: migrants who lived in the county where their new Hukou was registered, migrants who lived outside the Hukou registration county but within the newly registered prefecture city, migrants who lived outside the newly registered prefecture city, and migrants whose Hukou had been transferred to another prefecture city but still resided in their origin city (as shown in Figure D3). To exclude samples of the last case not captured by A2019e, we used the surveyed city followed by A2001.[4] Lastly, some people had migrated multiple times, and this situation can be complicated. They may have transferred their Hukous after moving to new prefecture cities, yet the cities where their Hukous were newly registered may not correspond to the variables that indicate when they moved. We specified these samples as migrants initially and then removed some of them if their destination cities and moving-in years were still mismatched after compiling.[5]

4 Recall that A2001 describes family relationships between respondents and interviewed persons.
5 We describe how we handled this situation in detail in the following section.

	hhid	pline	a2001	a2016b_short	a2019_short	a2019b	transfer	a2022m_short	surveyedCity
1	2017000273	2	2	<null>	3415	<null>	1	5107	3101
2	2017000301	2	2	<null>	4414	<null>	1	1408	4401
3	2017000355	2	2	<null>	5105	<null>	1	5002	5001
4	2017000356	2	2	<null>	4415	<null>	1	4416	4403
5	2017000490	2	10	<null>	6104	<null>	1	6110	6101
6	2017000512	1	1	<null>	3102	<null>	1	6529	3101
7	2017000530	2	2	<null>	1307	<null>	1	1310	1306
8	2017001441	1	1	<null>	4403	<null>	1	2201	2102
9	2017023822	2	10	<null>	3301	<null>	1	1101	3101
10	2017006173	1	1	<null>	2201	<null>	1	2205	2202
11	2017029672	5	5	<null>	1401	<null>	1	1401	1101
12	2017029672	3	3	<null>	1401	<null>	1	1401	1101
13	2015016963	2	2	<null>	3506	<null>	1	3501	3502
14	2013021805	2	2	<null>	4301	<null>	1	4303	5001
15	2015026797	2	2	<null>	3607	<null>	1	3601	4403
16	2017029672	4	5	<null>	1401	<null>	1	1401	1101
17	2017018396	2	2	<null>	3213	<null>	1	3209	3101
18	2013004508	1	2	<null>	1509	<null>	1	1307	1502
19	2015000846	5	10	<null>	4104	<null>	1	6542	1101

Figure 1D.4: Migrants with Hukou transfer: New and old Hukou registration cities and the surveyed city are all non-identical.

Information Pairwise

Now we successfully sorted out migrant observations, while the next issue lying ahead of us is to identify to which prefecture city each migrant moved, and accordingly, in which year they moved out and moved in. Foremost, variables A2016b, A2019, A2023j, and "surveyed city" all provide useful information. To apply the correct one, we need to further classify migrants into different situations. As always, we endeavored to improve the data accuracy and retain as many observations as possible.

```sql
case when migrant=1 and transfer=1 and a2023g=2 and a2016b is null and a2019_short!
=a2022m_short then a2019_short
        when migrant=1 and transfer=1 and a2023g=2 and a2016b is not null and a2016_revised!
        =a2022m_short then a2016_revised
        when migrant=1 and transfer=1 and a2023g=1 and a2023j is not null then a2023j
        when migrant=1 and transfer=1 and a2023g=1 and a2023j is null and a2019_short!
        =a2022m_short then a2019_short
        when migrant=1 and transfer=1 and a2023g is null and a2019e or a2019f is not null then
        a2016_revised
        when migrant=1 and transfer=1 and a2023g is null and (a2019e and a2019f) is null and
        a2019_short!=a2022m_short then a2019_short
        when migrant=1 and transfer=0 and a2023g is null and a2016b is null and hhead=1 and
        a2001=1 and surveyedCity!=substring(origin,4,4) then surveyedCity
        when migrant=1 and transfer=0 and a2023g is null and a2016b is not null then
        a2016_revised
        when migrant=1 and transfer=0 and a2023g=1 then a2023j else substring(origin,4,4)
        end destination
```

Figure 1D.5: SQL codes for identifying migration destination.

Here's the revised paragraph with these changes:

In general, the classification can be summarized as follows:

a. When A2023g equals 2, it suggests that individuals had never resided in other prefecture cities. While looking at migrants with Hukou transfer, sometimes the reference point that respondents took is the cities where their new Hukous were registered, instead of their "real" origin cities. Therefore, by setting transfer equals 1 and A2023g equals 2 followed by A2016b, we assigned either new Hukou registration cities or resident cities to those samples as their destination, depending on their specific situations.

b. When A2023g equals 1, it simply means that individuals were returnees. In general, A2023j tells us where individuals had migrated to before they returned to the Hukou registration cities. Yet for certain observations, this information is missing.

Instead, we assigned new Hukou registration cities as destination to some of them if they are migrants with Hukou transfer.

c. When A2023g is null, the CHFS annotates that this question was not assigned to those individuals because they lived in prefecture cities that their Hukous did not pertain to. In this case, we may learn about their resident cities using A2016b. However, sometimes even though we know in which city the migrant was living, his/her moving-out and/or moving-in year, referring to A2019e and A2019f, is still missing, and thus, these observations cannot be used. As an alternative solution to keeping information pairwise, we assigned the new Hukou registration cities to them and re-calculated their migration years by assuming they moved one year before they transferred their Hukous.[6]

d. Lastly, among the floating population, destination cities can either be A2016b or "surveyed city" depending on whether A2016 has a valid value. Again, we restricted applying the variable "surveyed city" to householders (hhead=1).

Besides the aforementioned, we assigned the origin cities (A2022m) to all natives as their destination. Moreover, as shown in Figure D6, sometimes the moving-out year (A2019e) is not the same as the moving-in year (A2019f) because individuals moved.

a2019b	a2019e	a2019f	transfer	a2022m_short	a2022l	a2023k	surveyedCity	leaveTime	arrivTime
<null>	<null>	<null>	1	6107	2016	<null>	5101	2015	2015
<null>	2001	2012	1	3301	2001	<null>	6101	2001	2012
<null>	1991	1991	1	5107	1980	<null> <1997	3101	<null>	<null>
<null>	<null>	<null>	1	1522	1979	1968	1201	<null>	<null>
<null>	2007	2007	1	1408	2014	<null>	4401	2007	2007
<null>	2013	2013	1	5326	2009	<null>	5323	2013	2013
<null>	2016	2016	1	5002	1966	<null>	5001	2016	2016
<null>	2007	2007	1	4416	2008	<null>	4403	2007	2007
<null>	<null>	2007	1	6110	2012	<null>	6101	2011	2007
<null>	<null>	<null>	1	6529	1964	1964	3101	<null>	<null>
<null>	<null>	<null>	1	4212	2008	<null>	3502	2007	2007
<null>	2009	2014	1	1310	2009	<null>	1306	2009	2014
<null>	<null>	<null>	1	3206	2015	<null>	3101	2014	2014
<null>	2014	2014	1	1401	2015	<null>	4331	2014	2014
<null>	<null>	<null>	1	1304	1999	<null>	1301	1998	1998
<null>	<null>	<null>	1	6109	1973	1970	6101	<null>	<null>
<null>	<null>	<null>	1	3611	1981	1971	3301	<null>	<null>
<null>	2015	2015	1	2310	2002	<null>	3101	2015	2015
<null>	2017	2017	1	3410	2017	<null>	3412	2017	2017

Figure 1D.6: Example of migration year.

Final Cleaning Procedure

Migrant Samples

Afterward, there were still some issues with the data because some observations had strange inputs. For example, the destination city did not always match the moving-in year, and a small group of respondents did not answer questions based on common knowledge, as shown in Figure D7. Dozens of samples had available inputs in A2023j, but the corresponding data that should have been stored in A2023k was missing. This caused the destination variable to extract data from A2023j but use a proxy (the year of Hukou transfer minus 1) as the migration time. To fix this problem, we could either

[6] In large cities, obtaining a local Hukou is quite hard for immigrants, especially since 2014, so it may take a much longer period for people to successfully transfer their Hukous. However, the number of such observations after cleaning is just 745 and among them, migrants who transferred their Hukous to Tier-1 cities are very few. We also confirm that the statistical significance of our trending indicator holds in the case of excluding those observations.

replace the destination variable with A2019 (the current Hukou registration city) or simply drop these observations. We chose the latter.

Additionally, some observations had identical inputs in A2022m and A2023j. This could be due to three possible reasons: the respondent took their current Hukou registration city as the reference point and considered where they came from as the place they had stayed in outside the Hukou registration city; individuals migrated and transferred the Hukous more than once; or individuals returned to their origin cities even after they transferred the Hukous. While we can easily identify the last group by setting A2022l<A2023k, it is difficult to distinguish the first two. As a result, although we know they are migrants, we cannot identify in which year these migrants moved to their resident cities. Therefore, we eventually dropped these observations as well.

a2016...	a2019...	a2019e	a2019f	tra...	a2022...	a2022l	a2023j	a2023k	leaveTime	arrivTime	destination
3205	3205	<null>	<null>	1	2303	2006	3717	<null>	2005	2005	3717
3205	3205	<null>	<null>	1	3701	2006	<null>	<null>	2005	2005	3205
2112	2112	<null>	<null>	1	2203	1994	<null>	<null>	<null>	<null>	2112
<null>	5002	<null>	<null>	1	5309	2016	5309	1985	2015	2015	5309
<null>	2201	<null>	<null>	1	3101	2001	3101	1997	1997	1997	3101
<null>	2201	<null>	<null>	1	2224	2012	2224	1989	2011	2011	2224
<null>	2201	<null>	<null>	1	3210	2008	3210	2003	2003	2003	3210
5101	5101	<null>	<null>	1	5002	1987	4403	1997	1997	1997	4403
<null>	4601	<null>	<null>	1	2102	2007	<null>	<null>	2006	2006	4601
4420	4290	2010	2010	1	4201	2015	<null>	<null>	2010	2010	4420
<null>	5002	<null>	<null>	1	5101	1987	5101	1974	<null>	<null>	5101
5309	5309	<null>	<null>	1	4502	2017	<null>	<null>	2016	2016	5309
<null>	1202	<null>	<null>	1	1410	2016	1410	1991	2015	2015	1410
<null>	1102	<null>	<null>	1	1507	2008	1507	2008	2008	2008	1507
<null>	2201	<null>	<null>	1	2224	2003	2224	1987	2002	2002	2224
<null>	2201	<null>	<null>	1	2204	1980	2204	1978	<null>	<null>	2204
<null>	3701	<null>	<null>	1	3412	2015	3702	2014	2014	2014	3702
6106	4401	2016	2016	1	6207	2000	<null>	<null>	2016	2016	6106
<null>	2201	<null>	<null>	1	2224	1998	<null>	<null>	1997	1997	2201
3302	3508	2012	2012	1	3505	2005	<null>	<null>	2012	2012	3302

Figure 1D.7: Remaining errors.

Recall that certain returnees may have migrated when they were quite young (under 16 years old). As our aim is to examine labor migration, teenagers are not suitable to be included in our analysis. After merging the datasets, we transformed the age variable to be time-variant. This enabled us to eliminate all the observations where the respondent's age was below 16 or above 65 between 1997 and 2017. In addition, for migrants who previously worked but did not have a job when they were interviewed in 2017, we matched regional employment statistics based on the industry categories of their last employment. Since migrants may have migrated after being unemployed or retired, we only kept respondents whose last job termination dates (A3139) were later than their migration time.

Native Samples

The cleaning process for native samples is much simpler. Here is a brief summary:

- Drop observations if A2022k (transfer) equals 1 but A2019 is null.
- Drop observations if A2022k (transfer) equals 1 but A2019b and A2023g are null.
- Drop observations if the city retrieved from the variable "surveyed city" differs from the origin city while the respondent was identified as a householder.

Lastly, among the full sample (both natives and migrants), observations were dropped if A3138 equals 2 (meaning they had never worked).

Variables Summary

Table 1D.1: Variables – Questions

Variables	*Questions*
A2001	XXX is your? [Example: Myself; Spouse; Parents; Children]
A2016b	In which province/city/county does XXX live?
A2019	Which province/city/county is the registered residence of XXX?
A2019b†	Is XXX Hukou registered in the villages/towns where he/she now lives?
A2019e‡	In which year did XXX leave [A2019]?
A2019f‡	In which year did XXX come to his/her resident province/city?
A2022k	Has XXX ever transferred the Hukou to another district/county?
A2022l	In which year did XXX experience his/her latest Hukou transfer?
A2022m	XXX's Hukou is moved out from which county/city/province?
A2023g∗	Has XXX ever left [A2019] for somewhere else for over 6 months?
A2023j∗	In which province/city did XXX live before returning?
A2023k∗	In which year did XXX go to [A2023j]?
A3138	Has XXX worked before?
A3139	When did XXX's last job end?

Notes: † denotes questions that are asked only if the CAPI system detects that the individual's resident county/district is identical to his/her Hukou registration county/district. ‡ denotes questions that are delivered only if the CAPI system detects that the individual's resident prefecture city is non-identical to his/her Hukou registration city. ∗ denotes questions that are asked only if the CAPI system detects that the resident prefecture city of an individual, whose age has been above 16 by the time of the survey, is identical to his/her current Hukou registration city.
Source: China Household Finance Survey (2017).

Appendix 2

Table 2A.1: Complete list of intra-city mobility clustering results

Cluster	City Name
1	Alar, Alxa League, Anqing, Baicheng, Baise, Baisha Li Autonomous County, Baoting Li and Miao Autonomous County, Bazhong, Bengbu, Binzhou, Bozhou, Changjiang Li Autonomous County, Changzhou, Chaoyang, Chizhou, Chuzhou, Dandong, Dingan County, Dongfang, Dongying, Fuyang, Fǔzhou, Ganzhou, Guigang, Hangzhou, Hanzhong, Hefei, Heyuan, Hezhou, Hinggan League, Huainan, Huaiyin, Huangshan, Huludao, Huzhou, Jian, Jiangmen, Jiaxing, Jilin, Jinan, Jingmen, Jiujiang, Jiyuan, Ledong Li Autonomous County, Liangshan Yi Autonomous Prefecture, Lianyungang, Lingao County, Lingshui Li Autonomous County, Liuan, Longyan, Luohe, Maanshan, Nanchang, Nanjing, Nantong, Ningbo, Ningde, Ordos, Panjin, Qianxinan Buyei and Miao Autonomous Prefecture, Qingdao, Qitaihe, Qujing, Quzhou, Rizhao, Shanghai, Shangrao, Shaoxing, Shenyang, Shizuishan, Songyuan, Suqian, Suzhou, Sùzhou, Taizhou, Tongliao, Tongling, Tàizhou, Weifang, Weihai, Wenchang, Wuhu, Wuxi, Wuzhou, Xilingol League, Xinyang Wenshan Zhuang and Miao Autonomous Prefecture, Xuancheng, Xuzhou, Yancheng, Yangjiang, Yangzhou, Yantai, Yibin, Yichang, Yingkou, Yulin, Yíchun, Zhenjiang, Zhoushan, Zibo, Zunyi, Changchun
2	Ankang, Anshan, Anshun, Anyang, Baoji, Bayannur, Bijie, Cangzhou, Changde, Changsha, Changzhi, Chengde, Chengdu, Chengmai County, Chenzhou, Chifeng, Chongzuo, Chuxiong Yi Autonomous Prefecture, Dali Bai Autonomous Prefecture, Dalian, Danzhou, Dazhou, Deyang, Dezhou, Enshi City, Fangchenggang, Foshan, Fuxin, Fuzhou, Guangan, Guangyuan, Guilin, Guyuan, Haikou, Handan, Harbin, Hebi, Hechi, Hegang, Hengshui, Hengyang, Heze, Honghe Hani and Yi Autonomous Prefecture, Huaibei, Huanggang, Huangshi, Huizhou, Hulunbuir, Jiaozuo, Jinchang, Jincheng, Jingdezhen, Jingzhou, Jinhua, Jining, Jixi, Kaifeng, Laibin, Leshan, Liaocheng, Liaoyang, Liaoyuan, Lijiang, Lincang, Linfen, Linyi, Lishui, Liupanshui, Liuzhou, Loudi, Luzhou, Maoming, Meishan, Meizhou, Mianyang, Nanchong, Nanning, Nanping, Nanyang, Panzhihua, Pingxiang, Puer, Putian, Puyang, Qiandongnan Miao and Dong Autonomous Prefecture, Qianjiang, Qiannan Buyei and Miao Autonomous Prefecture, Qingyuan, Qinzhou, Qionghai, Qiongzhong Li and Miao Autonomous County, Qiqihar, Quanzhou, Sanming, Sanya, Shangluo, Shangqiu, Shanwei, Shaoguan, Shaoyang, Shennongjia, Shuangyashan, Suining, Suizhou, Taian, Tangshan, Tianjin , Tianmen, Tongchuan, Tongren, Tunchang County, Ulanqab, Wanning, Weinan, Wenzhou, Wuhai, Wuzhong, Xiamen, Xi'an, Xiangfan, Xiangtan, Xiangxi Tujia and Miao Autonomous Prefecture, Xianning, Xiantao, Xianyang, Xiaogan, Xinxiang, Xinyu, Xishuangbanna Dai Autonomous Prefecture, Xuchang, Yaan, Yanan, Yanbian Korean Autonomous Prefecture, Yingtan, Yiyang, Yongzhou, Yueyang, Yunfu, Yùlin, Zaozhuang, Zhangjiajie, Zhangzhou, Zhanjiang, Zhaoqing, Zhaotong, Zhongshan, Zhoukou, Zhuhai, Zhumadian, Zhuzhou, Zigong, Ziyang

Table 2A.1: *Continued*

Cluster	City Name
3	Baishan, Baoding, Baoshan, Baotou, Beihai, Beijing , Benxi, Chaozhou, Chongqing, Da Hinggan Ling, Daqing, Dingxi, Diqing Tibetan Autonomous Prefecture, Ezhou, Fushun, Guiyang, Heihe, Huaihua, Jiamusi, Jieyang, Jinzhong, Jinzhou, Jiuquan, Kunming, Langfang, Luoyang, Lvliang, Mudanjiang, Neijiang, Ngawa Tibetan and Qiang Autonomous Prefecture, Pingdingshan, Qingyang, Qinhuangdao, Sanmenxia, Shantou, Shenzhen, Shijiazhuang, Shiyan, Shuozhou, Suihua, Taiyuan, Tieling, Tonghua, Tumxuk, Wuhan, Wuwei, Wuzhishan, Xingtai, Xinzhou, Yichun, Yuncheng, Yuxi, Zhangye, Zhongwei, Yinchuan
4	Aksu Prefecture, Baiyin, Datong, Dehong Dai and Jingpo Autonomous Prefecture, Dongguan, Gannan Tibetan Autonomous Prefecture, Garzê Tibetan Autonomous Prefecture, Guangzhou, Haibei Tibetan Autonomous Prefecture, Haidong, Hainan Tibetan Autonomous Prefecture, Haixi Mongol and Tibetan Autonomous Prefecture, Hami , Hohhot, Huangnan Tibetan Autonomous Prefecture, Kashgar Prefecture, Kizilsu Kyrgyz Autonomous Prefecture, Kokdala, Linxia Hui Autonomous Prefecture, Longnan, Nujiang Lisu Autonomous Prefecture, Pingliang, Sansha, Siping, Tianshui, Yangquan, Zhengzhou, Zhangjiakou
5	Alidiqu, Altay Prefecture, Bayingolin Mongol Autonomous Prefecture, Beitun, Bortala Mongol Autonomous Prefecture, Chamdo, Changji Hui Autonomous Prefecture, Golog Tibetan Autonomous Prefecture, Hotan Prefecture, Ili Kazakh Autonomous Prefecture, Jiayuguan, Karamay, Kunyu, Lanzhou, Macao, Nagqu, Shannan, Shigatse, Shuanghe, Tacheng Prefecture, Tiemenguan City, Turpan, Xining, Nyingchi
6	Hongkong, Lhasa, Shihezi, Urümqi, Wujiaqu, Yushu Tibetan Autonomous Prefecture

Source: Created by authors using Baidu Mobility Data.

Table 2A.2: Complete list of HW commuting clustering results

Cluster	City Name
Divergent	Changchun, Macao, Xi’an
1	Hongkong, Tiemenguan City, Shuanghe
2	Golog Tibetan Autonomous Prefecture, Kunyu, Kokdala
3	Chamdo, Garzê Tibetan Autonomous Prefecture, Haibei Tibetan Autonomous Prefecture, Hainan Tibetan Autonomous Prefecture, Hotan Prefecture, Huangnan Tibetan Autonomous Prefecture, Kizilsu Kyrgyz Autonomous Prefecture, Nagqu, Nujiang Lisu Autonomous Prefecture, Sansha, Tumxuk, Turpan

Table 2A.2: *Continued*

Cluster	City Name
4	Aksu Prefecture, Alar, Altay Prefecture, Alxa League, Baisha Li Autonomous County, Baoshan, Binzhou, Bortala Mongol Autonomous Prefecture, Changjiang Li Autonomous County, Chongzuo, Dingan County, Diqing Tibetan Autonomous Prefecture, Fangchenggang, Gannan Tibetan Autonomous Prefecture, Haixi Mongol and Tibetan Autonomous Prefecture, Hami , Kashgar Prefecture, Liangshan Yi Autonomous Prefecture, Lincang, Linxia Hui Autonomous Prefecture, Ngawa Tibetan and Qiang Autonomous Prefecture, Nyingchi, Puer, Qiongzhong Li and Miao Autonomous County, Shannan, Shigatse, Tacheng Prefecture, Tunchang County, Wenchang, Wujiaqu, Zhaotong
5	Baise, Baoting Li and Miao Autonomous County, Bayannur, Bayingolin Mongol Autonomous Prefecture, Bijie, Cangzhou, Changji Hui Autonomous Prefecture, Chaoyang, Chengmai County, Chuxiong Yi Autonomous Prefecture, Chuzhou, Da Hinggan Ling, Dali Bai Autonomous Prefecture, Dandong, Danzhou, Dehong Dai and Jingpo Autonomous Prefecture, Dezhou, Dingxi, Dongying, Guigang, Guyuan, Haidong, Hechi, Hengshui, Heze, Hezhou, Huludao, Ili Kazakh Autonomous Prefecture, Jian, Jinchang, Jincheng, Jining, Jiuquan, Karamay, Laibin, Ledong Li Autonomous County, Liaocheng, Lijiang, Lingao County, Linyi, Longnan, Nantong, Ordos, Panzhihua, Qinzhou, Qionghai, Quzhou, Rizhao, Sùzhou, Taian, Tangshan, Tàizhou, Wanning, Weifang, Wenshan Zhuang and Miao Autonomous Prefecture, Wuwei, Wuzhong, Xingtai, Xuancheng, Yantai, Yunfu, Yushu Tibetan Autonomous Prefecture, Yíchun, Yùlin, Zhangye, Zhongwei, Zibo, Ziyang
6	Anqing, Baiyin, Bengbu, Bozhou, Changzhi, Changzhou, Chaozhou, Chengde, Chifeng, Chizhou, Deyang, Dongfang, Fuyang, Fǔzhou, Handan, Hanzhong, Hinggan League, Honghe Hani and Yi Autonomous Prefecture, Huaibei, Huaihua, Huainan, Huaiyin, Huanggang, Huangshi, Hulunbuir, Huzhou, Jiangmen, Jiaozuo, Jiaxing, Jiayuguan, Jingmen, Jinzhong, Jinzhou, Jiyuan, Leshan, Lianyungang, Liaoyang, Liaoyuan, Linfen, Liuan, Liupanshui, Longyan, Lvliang, Maanshan, Maoming, Meizhou, Meishan, Nanping, Neijiang, Ningde, Panjin, Pingliang, Pingxiang, Qianjiang, Qiannan Buyei and Miao Autonomous Prefecture, Qianxinan Buyei and Miao Autonomous Prefecture, Qingyang, Qinhuangdao, Qujing, Shangqiu, Shangrao, Shizuishan, Shuozhou, Suizhou, Suqian, Suzhou, Tianmen, Tianshui, Tongliao, Tongling, Tongren, Ulanqab, Urümqi, Weihai, Wuhai, Wuhu, Wuzhishan, Wuzhou, Xinxiang, Xinzhou, Xishuangbanna Dai Autonomous Prefecture, Xuzhou, Yaan, Yancheng, Yangjiang, Yangzhou, Yibin, Yingkou, Yingtan, Yulin, Yuncheng, Yuxi, Zhangzhou, Zhanjiang, Zhaoqing, Zhenjiang, Zhumadian, Zigong, Zaozhuang

Table 2A.2: *Continued*

Cluster	City Name
7	Alidiqu, Ankang, Anshan, Anshun, Anyang, Baicheng, Baoding, Baoji, Baotou, Bazhong, Beihai, Beitun, Benxi, Changde, Chenzhou, Dalian, Daqing, Datong, Dazhou, Enshi City, Ezhou, Fushun, Fuxin, Ganzhou, Guangan, Guangyuan, Guilin, Hebi, Hegang, Heyuan, Huangshan, Jieyang, Jilin, Jinan, Jingdezhen, Jingzhou, Jiujiang, Jixi, Kaifeng, Langfang, Lingshui Li Autonomous County, Lishui, Liuzhou, Loudi, Luohe, Luzhou, Mianyang, Nanning, Nanyang, Ningbo, Pingdingshan, Putian, Qiandongnan Miao and Dong Autonomous Prefecture, Qingdao, Qingyuan, Qiqihar, Sanmenxia, Sanming, Shangluo, Shantou, Shanwei, Shaoguan, Shaoxing, Shaoyang, Shihezi, Shijiazhuang, Shiyan, Shuangyashan, Siping, Songyuan, Suining, Taizhou, Tianjin , Tieling, Tongchuan, Tonghua, Weinan, Wuxi, Xiangfan, Xiangxi Tujia and Miao Autonomous Prefecture, Xiantao, Xianyang, Xiaogan, Xilingol League, Xinyang, Xinyu, Xuchang, Yanan, Yangquan, Yichang, Yinchuan, Yiyang, Yongzhou, Yueyang, Zhangjiajie, Zhangjiakou, Zhoukou, Zhoushan, Zunyi, Zhuzhou, Puyang
8	Baishan, Beijing , Changsha, Chengdu, Chongqing, Foshan, Fuzhou, Guangzhou, Guiyang, Haikou, Harbin, Hefei, Heihe, Hengyang, Hohhot, Huizhou, Jiamusi, Jinhua, Kunming, Lanzhou, Lhasa, Luoyang, Mudanjiang, Nanchang, Nanchong, Nanjing, Qitaihe, Quanzhou, Sanya, Shenyang, Suihua, Taiyuan, Wenzhou, Wuhan, Xiamen, Xiangtan, Xianning, Xining, Yanbian Korean Autonomous Prefecture, Yichun, Zhengzhou, Zhongshan, Zhuhai,Hangzhou
9	Dongguan, Shanghai, Shennongjia, Shenzhen

Source: Created by authors using Baidu Mobility Data.

Table 2A.3: Complete list of DLR travel clustering results

Cluster	City Name
Divergent	Hongkong, Qinzhou, Shennongjia, Songyuan
1	Alidiqu, Altay Prefecture, Beitun, Chaozhou, Chizhou, Da Hinggan Ling, Dali Bai Autonomous Prefecture, Dehong Dai and Jingpo Autonomous Prefecture, Diqing Tibetan Autonomous Prefecture, Dongguan, Foshan, Fuzhou, Guangyuan, Guangzhou, Guiyang, Haikou, Heyuan, Huizhou, Jiangmen, Jieyang, Jinhua, Kunming, Lhasa, Lingshui Li Autonomous County, Luoyang, Nanjing, Ngawa Tibetan and Qiang Autonomous Prefecture, Nyingchi, Putian, Sanming, Sansha, Sanya, Shannan, Shantou, Shenzhen, Shiyan, Taizhou, Wanning, Wenchang, Wenzhou, Xiamen, Xiangxi Tujia and Miao Autonomous Prefecture, Xishuangbanna Dai Autonomous Prefecture, Yaan, Yangjiang, Yulin, Yunfu, Yushu Tibetan Autonomous Prefecture, Yuxi, Zhongshan, Zhoushan, Zhuhai, Garzê Tibetan Autonomous Prefecture

Table 2A.3: *Continued*

Cluster	City Name
2	Ankang, Anqing, Anshun, Anyang, Baoding, Baoji, Baoshan, Baoting Li and Miao Autonomous County, Baotou, Bazhong, Beihai, Beijing , Bijie, Bortala Mongol Autonomous Prefecture, Chamdo, Changde, Changjiang Li Autonomous County, Changsha, Changzhi, Changzhou, Chengdu, Chengmai County, Chenzhou, Chongqing, Chuxiong Yi Autonomous Prefecture, Chuzhou, Dalian, Danzhou, Daqing, Datong, Dazhou, Deyang, Dingan County, Dongfang, Dongying, Enshi City, Ezhou, Fushun, Fuyang, Fǔzhou, Gannan Tibetan Autonomous Prefecture, Ganzhou, Golog Tibetan Autonomous Prefecture, Guigang, Guilin, Haibei Tibetan Autonomous Prefecture, Hainan Tibetan Autonomous Prefecture, Haixi Mongol and Tibetan Autonomous Prefecture, Hami , Hangzhou, Hanzhong, Hebi, Hechi, Hefei, Hegang, Heihe, Hengyang, Hezhou, Honghe Hani and Yi Autonomous Prefecture, Huaibei, Huaihua, Huaiyin, Huanggang, Huangnan Tibetan Autonomous Prefecture, Huangshan, Huangshi, Huzhou, Ili Kazakh Autonomous Prefecture, Jiamusi, Jiaozuo, Jinan, Jinchang, Jincheng, Jingdezhen, Jingmen, Jingzhou, Jining, Jiujiang, Jiuquan, Jixi, Jiyuan, Kaifeng, Karamay, Kashgar Prefecture, Lanzhou, Ledong Li Autonomous County, Leshan, Liangshan Yi Autonomous Prefecture, Lianyungang, Lijiang, Linfen, Lingao County, Linxia Hui Autonomous Prefecture, Lishui, Liupanshui, Liuzhou, Longnan, Longyan, Loudi, Luzhou, Lvliang, Maanshan, Maoming, Meishan, Meizhou, Mianyang, Nagqu, Nanchong, Nanning, Nanping, Nanyang, Neijiang, Ningbo, Ningde, Ordos, Panzhihua, Pingdingshan, Pingliang, Pingxiang, Puer, Puyang, Qiandongnan Miao and Dong Autonomous Prefecture, Qianjiang, Qiannan Buyei and Miao Autonomous Prefecture, Qianxinan Buyei and Miao Autonomous Prefecture, Qingdao, Qingyang, Qingyuan, Qionghai, Quanzhou, Qujing, Quzhou, Sanmenxia, Shangluo, Shanwei, Shaoguan, Shaoxing, Shaoyang, Shihezi, Shijiazhuang, Suining, Suizhou, Tacheng Prefecture, Taiyuan, Tianmen, Tongchuan, Tongling, Tongren, Tàizhou, Urümqi, Weinan, Wuhai, Wuhan, Wuxi, Wuzhishan, Wuzhou, Xian, Xiangfan, Xiangtan, Xianning, Xiantao, Xianyang, Xiaogan, Xilingol League, Xingtai, Xinxiang, Xinyang, Xinzhou, Xuancheng, Xuchang, Yanan, Yangquan, Yangzhou, Yibin, Yichang, Yinchuan, Yingtan, Yiyang, Yongzhou, Yueyang, Yuncheng, Yíchun, Yùlin, Zaozhuang, Zhangjiajie, Zhangye, Zhangzhou, Zhanjiang, Zhaoqing, Zhengzhou, Zhenjiang, Zhoukou, Zhumadian, Zhuzhou, Zigong, Zunyi
3	Aksu Prefecture, Anshan, Baise, Baishan, Baiyin, Bayannur, Bayingolin Mongol Autonomous Prefecture, Bengbu, Benxi, Bozhou, Cangzhou, Changji Hui Autonomous Prefecture, Dezhou, Dingxi, Fangchenggang, Fuxin, Guangan, Guyuan, Haidong, Handan, Harbin, Hengshui, Heze, Hinggan League, Hohhot, Hotan Prefecture, Hulunbuir, Jian, Jiaxing, Jiayuguan, Jinzhong, Jinzhou, Kizilsu Kyrgyz Autonomous Prefecture, Kunyu, Laibin, Langfang, Liaocheng, Liaoyang, Linyi, Liuan, Luohe, Mudanjiang, Nanchang, Nantong, Nujiang Lisu Autonomous Prefecture, Panjin, Qiqihar, Qitaihe, Rizhao, Shanghai , Shangqiu, Shangrao, Shenyang, Shizuishan, Shuangyashan, Shuozhou, Suqian, Sùzhou, Taian, Tianshui, Tiemenguan City, Tumxuk, Tunchang County, Turpan, Ulanqab, Weifang, Weihai, Wenshan Zhuang and Miao Autonomous Prefecture, Wuhu, Wuwei, Wuzhong, Xining, Xinyu, Xuzhou, Yanbian Korean Autonomous Prefecture, Yancheng, Yantai, Yichun, Zhangjiakou, Zhaotong, Zhongwei, Zibo, Ziyang, Shigatse

Table 2A.3: *Continued*

Cluster	City Name
4	Alar, Alxa League, Baisha Li Autonomous County, Binzhou, Changchun, Chaoyang, Chengde, Chifeng, Huainan, Lincang, Qinhuangdao, Suihua, Suzhou, Tangshan, Tianjin, Tonghua, Tongliao, Wujiaqu
5	Baicheng, Chongzuo, Dandong, Huludao, Jilin, Liaoyuan, Qiongzhong Li and Miao Autonomous County, Siping, Tieling, Yingkou
6	Kokdala, Macao, Shuanghe

Source: Created by authors using Baidu Mobility Data.

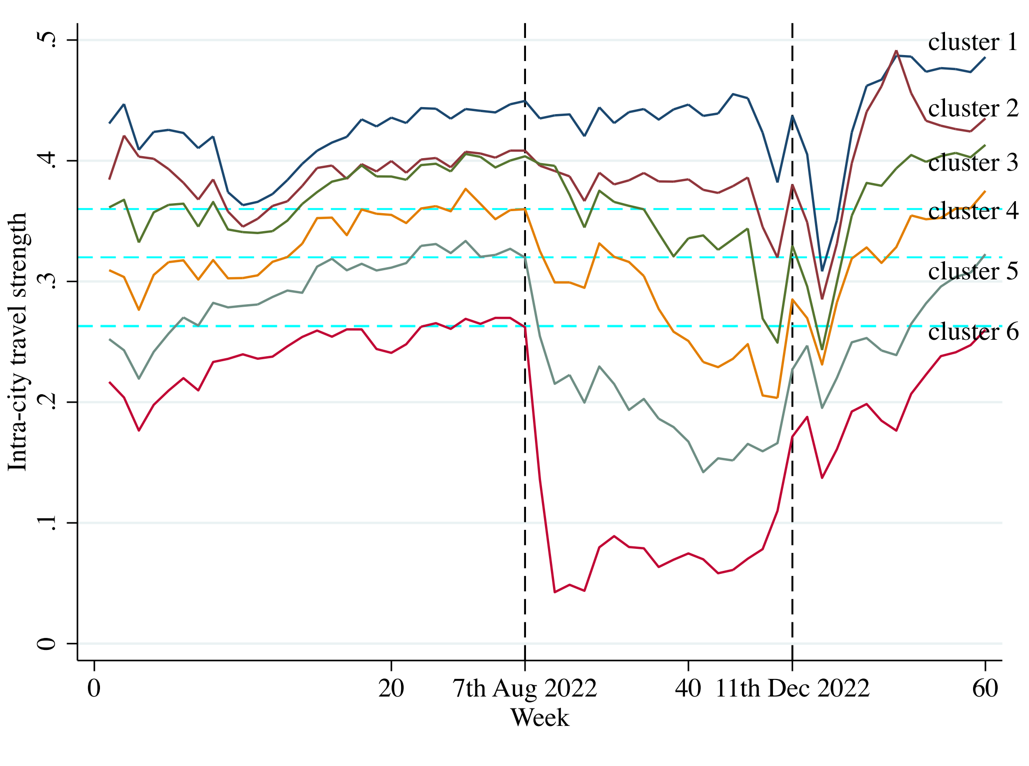


Figure 2A.1: Data visualization for the average intra-city travel strength of each cluster.

Source: Created by authors using the Baidu Mobility Data.

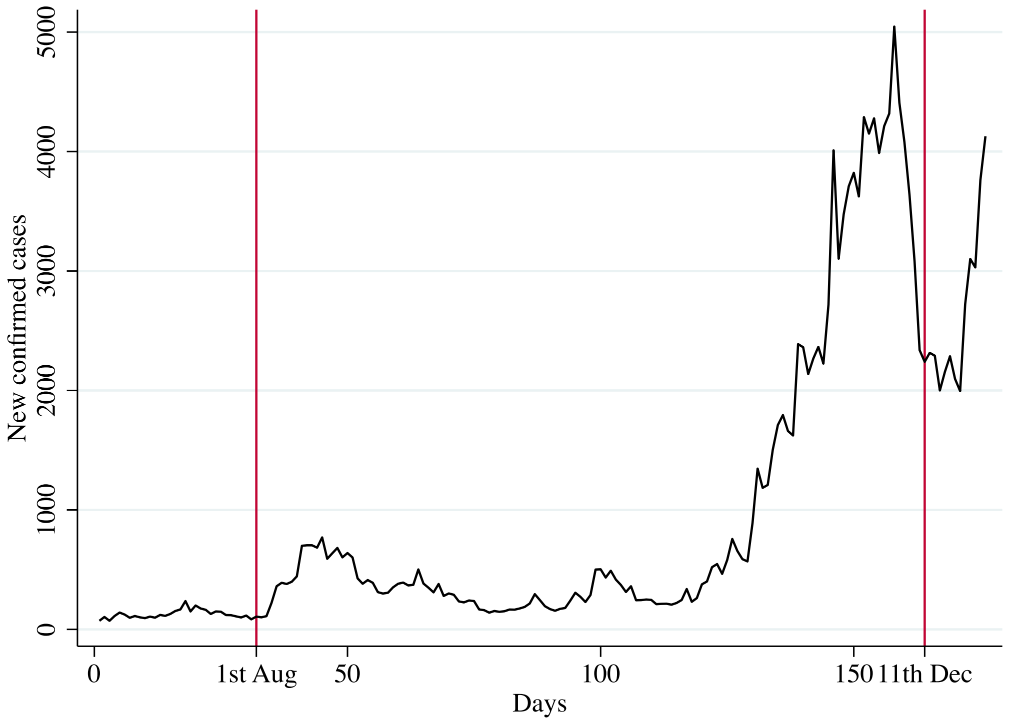


Figure 2A.2: Daily new confirmed COVID-19 cases in China during July 1st–December 23rd, 2022.

Source: National Health Commission of the PRC.

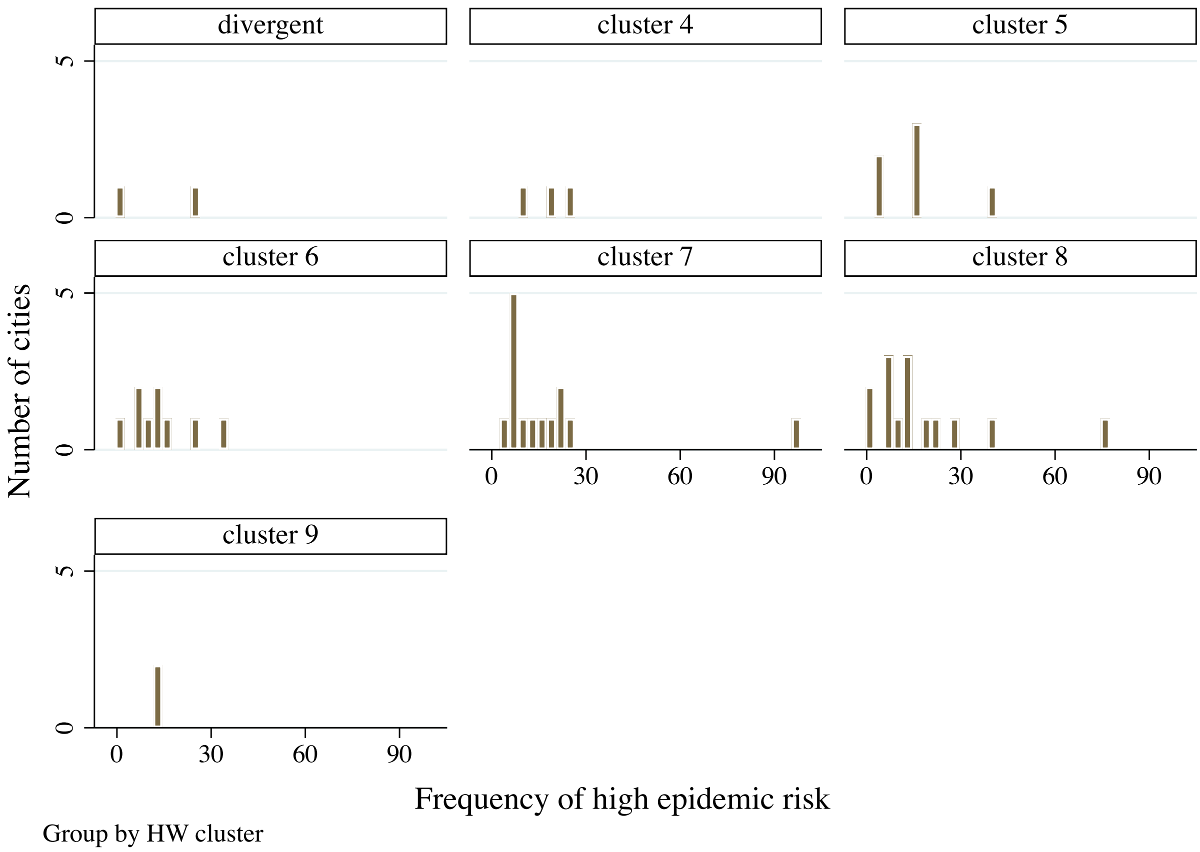


Figure 2A.3: Frequency of exposure to high COVID-19 risk grouped by clustering results of HW commuting.

Source: Created by authors using the Baidu Mobility Data.

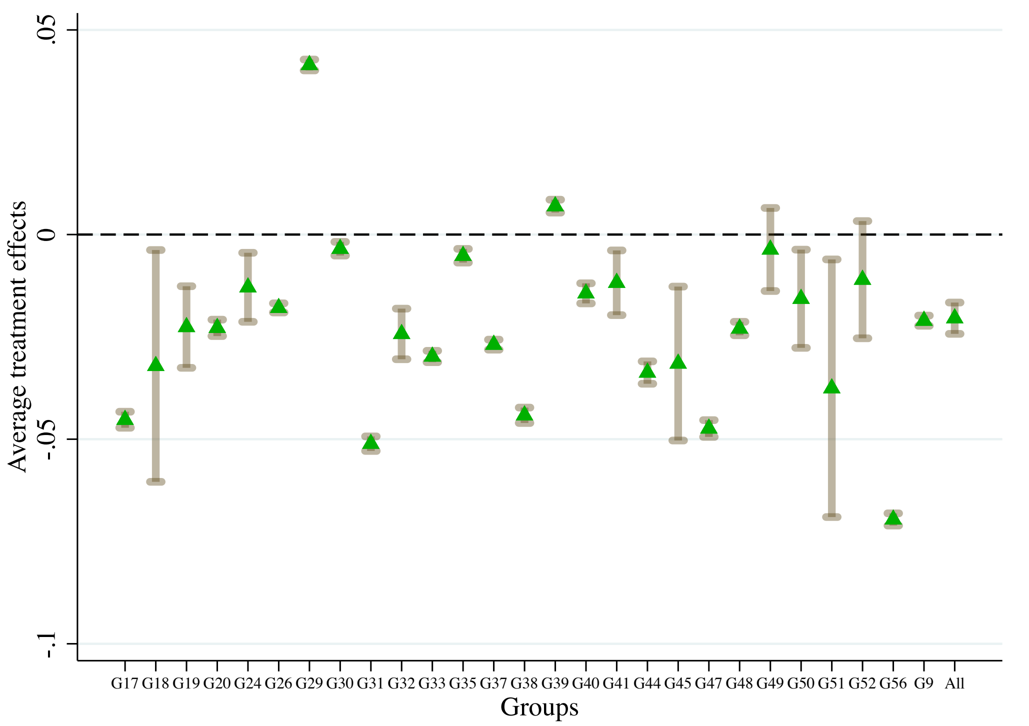


Figure 2A.4: Partially aggregated group-specific effects.

Source: Created by authors using the Baidu Mobility Data and COVID-19 risk-level data (Gong et al. 2023).

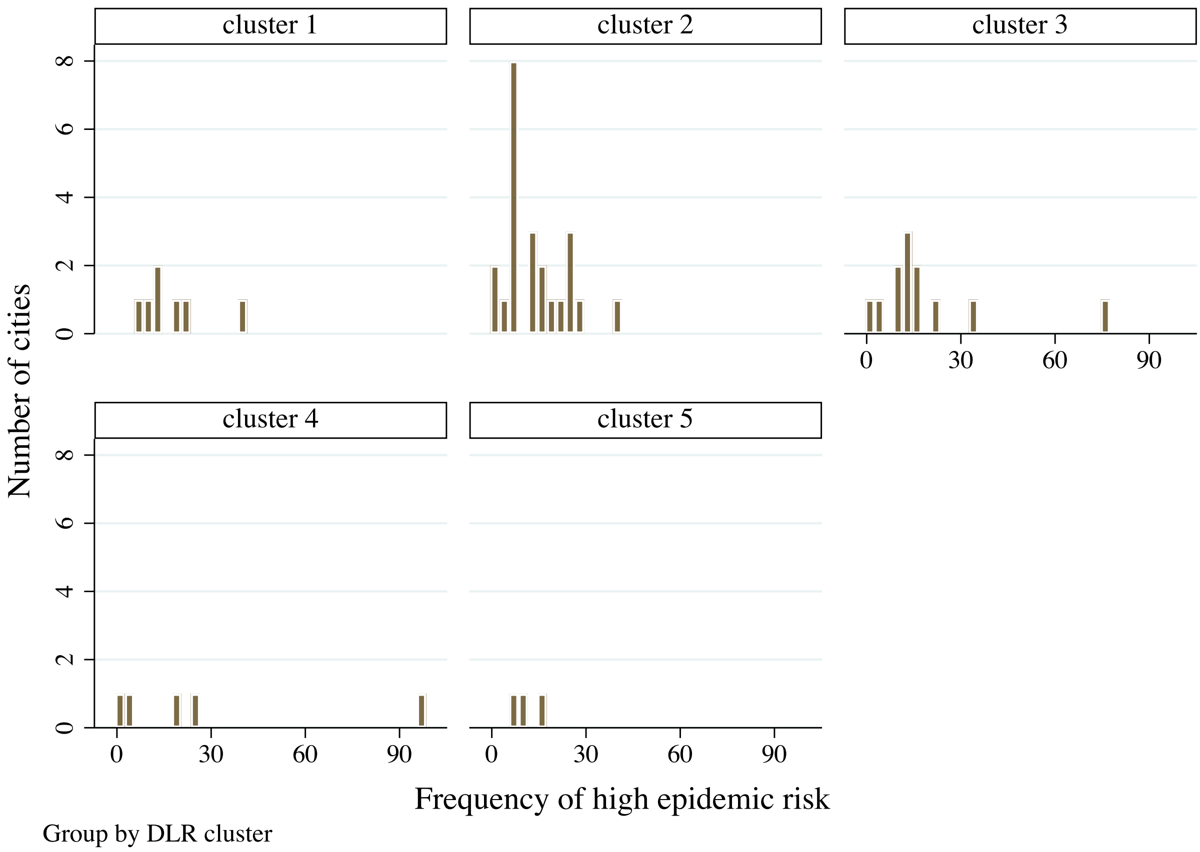


Figure 2A.5: Frequency of exposure to high COVID-19 risk grouped by clustering results of DLR travel.

Source: Created by authors using the Baidu Mobility Data.

Appendix 3

Table 3A.1: Questions About Test Scores, SHD, Parental Practices & Absence of Parenting

	Content
Test Scores	
Rate: poor, average, good, very good	
Chinese	As far as you know, what was the child's average grade in Chinese language or grammar last semester?
Mathematics	As far as you know, what was the child's average grade in math last semester?
Study Habits & discipline	
Rate: strongly disagree, disagree, agree, strongly agree	
QA1	This child studies very hard.
QA2	When this child finishes his/her homework, he/she checks it many times to see if he/she did it correctly.
QA3	This child plays only after he/she finished his/her homework.
QA4	During class-time, this child is concentrated on the things he/she does.
QA5	This child respects the rules and the order.
QA6	Once he/she starts to do something, this child will complete it no matter what happens.
QA7	This child likes to keep all his/her school things in great order.
Parental Practices	
	By parents themselves
Rate: never, rarely (once a month), sometimes (once a week), often (2-4 times a week), very often (5-7 times a week)	
QB1	How often did you give up watching TV shows you liked to avoid disturbing your child when he/she was studying?
QB2	How often have you discussed what happens at school with your child since this semester/last semester?
QB3	How often did you ask the child to finish homework this semester/last semester?
QB4	How often did you check the child's homework this semester/last semester?
QB5	How often did you restrict or stop the child from watching TV this semester/last semester?
QB6	How often did you restrict certain types of TV programs the child could watch this semester/last semester?
	By the interviewer
Rate: strongly disagree, disagree, neither agree nor disagree, agree, strongly agree	
QB7	Home environment (such as child's artwork, books, or other study materials) indicates that the parents care about the child's education.
QB8	The parents take the initiative to actively communicate with the child.
Absence of Parenting	
If the answers are neither of the parents in QC1 and QC2 and 0 in QC3, then the dummy "absence" is coded as 1, and 0 otherwise.	
QC1	Who mainly takes care of the child at daytime?
QC2	Who mainly takes care of the child at night?
QC3	How many times could the child meet his/her parent(s) per week on average?

Notes: Parental practices are measured with the geometric mean of the average normalized outcome of QB1–QB6 and QB7–QB8, respectively. We coded the answers into numeric values and assigned higher figures to more positive evaluations. English translation is in terms of the CFPS 2018 questionnaire.
Source: CFPS data.

Table 3A.2: Descriptive Statistics – Supplementary Spending Priority Ranking

Variable	Description	Obs.	Mean	Std. Dev.	Min	Max
diet	Spending priority for food and drink including eating out	12,665	9.37	1.11	1	10
transport	Spending priority for local transportation and post, and telecommunications	12,665	6.70	1.87	1	10
utility	Spending priority for housing mortgage, rent, utilities and property management	12,665	6.51	2.05	1	10
necessities	Spending priority for daily necessities, home repairs, cars, other transport tools, furnitures and electrical appliances	12,665	6.58	2.14	1	10
healthcare	Spending priority for medication, healthcare and fitness	12,665	5.36	2.29	1	10
education	Spending priority for education, culture and recreation, and travel	12,665	6.73	2.09	1	10
donation	Spending priority for financial support given to others and social donation	12,665	3.55	1.60	1	10
insurance	Spending priority for business insurance	12,665	3.64	1.82	1	10
clothing	Spending priority for clothes and beauty (e.g., haircut, spa, cosmetics)	12,665	6.09	1.59	1	10
other	Spending priority for all other items	12,665	4.91	1.59	1	10

Notes: The amount paid for some items was asked on a monthly basis. For calculation purposes, we converted them into annual quantities. The larger the value, the higher the priority. All statistics were adjusted using the sampling weights.
Source: Authors' elaboration using CFPS data.

Table 3A.3: Estimates Using Common-Factor-Analysis Based CADI

	(1) Baseline	(2) Main	(3) Main Lag	(4) One-Child	(5) Multi-Child
PAI	0.1077***	0.1289***	0.1120***	0.1728***	0.0866***
	(0.0176)	(0.0216)	(0.0279)	(0.0287)	(0.0292)
diet	0.0048***	0.0056***	0.0021	0.0047**	0.0061***
	(0.0013)	(0.0014)	(0.0018)	(0.0022)	(0.0020)
transport	0.0073***	0.0064***	0.0050***	0.0069***	0.0060***
	(0.0010)	(0.0012)	(0.0017)	(0.0018)	(0.0016)
utility	0.0048***	0.0047***	0.0047***	0.0074***	0.0019
	(0.0011)	(0.0013)	(0.0016)	(0.0019)	(0.0017)
necessities	0.0031***	0.0040***	0.0040**	0.0045**	0.0030*
	(0.0010)	(0.0012)	(0.0016)	(0.0018)	(0.0016)
healthcare	0.0076***	0.0067***	0.0038**	0.0080***	0.0053***
	(0.0010)	(0.0011)	(0.0018)	(0.0019)	(0.0015)
education	0.0055***	0.0061***	0.0041**	0.0050***	0.0070***
	(0.0011)	(0.0012)	(0.0018)	(0.0019)	(0.0015)
donation	0.0049***	0.0044***	-0.0000	0.0035	0.0050***
	(0.0013)	(0.0014)	(0.0019)	(0.0022)	(0.0018)
insurance	0.0003	0.0008	0.0020	0.0025	-0.0014
	(0.0011)	(0.0012)	(0.0018)	(0.0020)	(0.0018)
clothing	0.0039***	0.0040***	0.0060***	0.0037*	0.0041**
	(0.0012)	(0.0013)	(0.0018)	(0.0022)	(0.0016)
other	0.0022*	0.0023*	0.0003	0.0030	0.0015
	(0.0012)	(0.0012)	(0.0017)	(0.0019)	(0.0017)
parent_age		0.0025***	0.0033***	0.0036***	0.0013*
		(0.0005)	(0.0007)	(0.0006)	(0.0007)
child gender		-0.0414***	-0.0453***	-0.0442***	-0.0376***
		(0.0041)	(0.0057)	(0.0055)	(0.0059)
sleep		0.0081***	0.0103***	0.0072**	0.0076**
		(0.0024)	(0.0031)	(0.0031)	(0.0032)
child health		-0.0021**	-0.0023**	-0.0041***	-0.0011
		(0.0009)	(0.0011)	(0.0012)	(0.0012)
edu_savings		0.0155***	0.0169***	0.0139**	0.0209***
		(0.0046)	(0.0063)	(0.0065)	(0.0069)
marital		-0.0290**	-0.0343**	-0.0389***	-0.0119
		(0.0116)	(0.0174)	(0.0139)	(0.0184)
parental practices		0.0536***	0.0234	0.0016	0.0985***
		(0.0132)	(0.0179)	(0.0195)	(0.0185)
residence		-0.0236***	-0.0228***	-0.0210***	-0.0231***
		(0.0047)	(0.0064)	(0.0064)	(0.0063)
absence		0.0364***	0.0237	0.0375**	0.0328**
		(0.0101)	(0.0157)	(0.0146)	(0.0135)
key_school		0.0216***	0.0181***	0.0182***	0.0246***
		(0.0045)	(0.0056)	(0.0063)	(0.0064)
reaction (passive)		0.0279*	0.0357*	0.0615***	0.0047
		(0.0160)	(0.0195)	(0.0228)	(0.0233)
reaction (positive)		0.0541***	0.0592***	0.0812***	0.0346***
		(0.0064)	(0.0085)	(0.0106)	(0.0075)
Constant	0.6314***	0.4052***	0.3733***	0.3752***	0.4432***
	(0.0023)	(0.0309)	(0.0421)	(0.0393)	(0.0448)
Survey wave FE	N	Y	Y	Y	Y
Province FE	N	Y	Y	Y	Y
Birth cohort FE	N	Y	Y	Y	Y
Num. of children FE	N	Y	Y	N	Y
Obs	10299	7811	4072	3685	4125
R squared	0.0121	0.0895	0.0974	0.1210	0.0899

Notes: The CADI is normalized. The PAI and SPR variables are centred. In Column (3), all SPR variables are lagged by two periods because the CFPS was conducted biennially. Robust standard errors shown in parentheses are clustered at the province-birth cohort level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Source: Created by authors using CFPS data.

Table 3A.4: Spending Priorities, Parent Advantages and Child Academic Development

	(1) Score	(2) SHD	(3) Score (new SPR)	(4) SHD (new SPR)
PAI	0.2650***	0.1078***	0.3003***	0.1113***
	(0.0377)	(0.0212)	(0.0426)	(0.0231)
diet	-0.0014	0.0060***	0.0024	0.0041
	(0.0023)	(0.0014)	(0.0043)	(0.0025)
transport	-0.0022	0.0069***	0.0006	0.0079***
	(0.0021)	(0.0012)	(0.0032)	(0.0017)
utility	-0.0042**	0.0056***	-0.0003	0.0041**
	(0.0020)	(0.0013)	(0.0027)	(0.0018)
necessities	-0.0034*	0.0048***	-0.0017	0.0029*
	(0.0019)	(0.0012)	(0.0027)	(0.0017)
healthcare	-0.0005	0.0072***	0.0019	0.0054***
	(0.0018)	(0.0012)	(0.0027)	(0.0016)
education	0.0056***	0.0057***	0.0069**	0.0041**
	(0.0019)	(0.0012)	(0.0029)	(0.0018)
donation	0.0010	0.0046***	0.0028	0.0047**
	(0.0022)	(0.0015)	(0.0034)	(0.0021)
insurance	-0.0032	0.0014	-0.0013	0.0001
	(0.0021)	(0.0012)	(0.0029)	(0.0019)
clothing	0.0034	0.0039***	0.0019	0.0024
	(0.0022)	(0.0013)	(0.0033)	(0.0020)
other	-0.0018	0.0026**	(omitted)	(omitted)
	(0.0020)	(0.0013)		
parent_age	0.0012	0.0025***	0.0012	0.0029***
	(0.0008)	(0.0005)	(0.0009)	(0.0006)
child gender	-0.0537***	-0.0379***	-0.0576***	-0.0388***
	(0.0072)	(0.0039)	(0.0083)	(0.0048)
sleep	-0.0061	0.0090***	-0.0065	0.0100***
	(0.0039)	(0.0024)	(0.0044)	(0.0028)
child health	-0.0049***	-0.0017**	-0.0048***	-0.0021**
	(0.0014)	(0.0008)	(0.0015)	(0.0009)
edu_savings	0.0351***	0.0120**	0.0330***	0.0127**
	(0.0077)	(0.0046)	(0.0094)	(0.0058)
marital	-0.0316*	-0.0258**	-0.0603***	-0.0188
	(0.0182)	(0.0117)	(0.0200)	(0.0141)
parental practices	0.1971***	0.0396***	0.1918***	0.0169
	(0.0231)	(0.0133)	(0.0270)	(0.0161)
residence	0.0344***	-0.0301***	0.0322***	-0.0329***
	(0.0081)	(0.0047)	(0.0101)	(0.0058)
absence	0.0255	0.0361***	0.0110	0.0346***
	(0.0189)	(0.0096)	(0.0192)	(0.0121)
key_school	0.0383***	0.0179***	0.0273***	0.0184***
	(0.0078)	(0.0045)	(0.0098)	(0.0050)
reaction (passive)	0.0620***	0.0230	0.0572*	0.0291
	(0.0237)	(0.0156)	(0.0298)	(0.0177)
reaction (positive)	0.0837***	0.0467***	0.0762***	0.0538***
	(0.0108)	(0.0064)	(0.0140)	(0.0073)
Constant	0.4114***	0.4307***	0.4379***	0.4162***
	(0.0566)	(0.0301)	(0.0622)	(0.0384)
Time FE	Y	Y	Y	Y
Province FE	Y	Y	Y	Y
Birth FE	Y	Y	Y	Y
Num. of child. FE	Y	Y	Y	Y
Obs	8742	7846	5774	5206
R squared	0.1293	0.0897	0.1260	0.0971

Notes: The PAI and SPR variables are centred. DVs are normalized. Robust standard errors shown in parentheses are clustered at the province-birth cohort level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Source: Created by authors using CFPS data.