7-2022

Externalities in the housing market and agglomeration economies

Yifan WU
Singapore Management University, yifan.wu.2017@phdecons.smu.edu.sg

EXTERNALITIES IN THE HOUSING MARKET AND AGGLOMERATION ECONOMIES

By

YIFAN WU

A DISSERTATION

In

ECONOMICS

Presented to the Singapore Management University in Partial Fulfilment

of the Requirements for the Degree of PhD in Economics

2022

Supervisor of Dissertation

PhD in Economics, Programme Director

EXTERNALITIES IN THE HOUSING MARKET AND AGGLOMERATION ECONOMIES

by
Yifan Wu

Submitted to the School of Economics in Partial Fulfilment of the
Requirements for the Degree of Doctor of Philosophy in Economics

Dissertation Committee:

Zhenlin Yang (Supervisor/Chair)
Professor of Economics & Statistics
Singapore Management University

Jing Li (Co-chair)
Assistant Professor of Economics
Singapore Management University

Jungho Lee
Assistant Professor of Economics
Singapore Management University

Ying Deng
Associate Professor of Economics
University of International Business and Economics

Singapore Management University
2022

Abstract

This thesis studies the externalities in the housing market and agglomeration economies. While knowledge-based externalities, or knowledge spillovers are one of the most important micro-foundations of agglomeration economies, the first chapter studies how knowledge spillovers from universities affect local innovation activities. In the second chapter, we propose a high-order spatiotemporal autoregression approach to study the externalities in the housing market. The third chapter studies another important but under explored aspect of the agglomeration economies – the role that marriage market plays in providing incentives to promote urbanization, along with the unique feminization phenomenon during this process.

The first chapter studies the impact of universities on local innovation activity by exploiting a unique university expansion policy in China as a quasi-experiment. In this chapter, we take a geographic approach, empowered by geocoded data on patents and new products at the address level, to identify knowledge spillovers as an important channel. We obtain three main findings. First, university expansion significantly increases universities' own innovation capacity, which results in a dramatic boom of local industry patents. Second, the impact of university expansion on local innovation activities attenuates sharply within 2 kilometers of the universities. Third, university expansion boosts nearby firms' new products and the number of nearby industrial patents that cite university patents but not industry patents that cite patents far away from universities.

In the second chapter, we propose a high-order spatiotemporal autoregression approach for analyzing large real estate prices data. Real estate prices arrive sequentially on different housing units over time in a large volume. In this paper, we propose a high-order spatiotemporal autoregressive model with unobserved cluster and time heterogeneity. When the numbers of clusters (C) and time segments

(T) are finite and the errors are iid, quasi maximum likelihood method is used for model estimation and inference. In the presence of unknown heteroskedasticity, or C and/or T is large, an *adjusted quasi score* method is proposed for model estimation and inference. Methods for constructing the space-time connectivity matrices are proposed. Monte Carlo experiments are performed for assessing the finite sample properties of the proposed methods. An empirical application is presented using the housing transaction data in Beijing. We find that the estimation of the spatiotemporal interaction effects are largely affected after controlling for cluster heterogeneity at the community level.

The third chapter studies the relationship between urbanization and feminization, where the marriage market plays an important role in connecting the two. Previous literature studying urbanization and migration has mainly considered incentives arising from cross-city variation in productivity and the subsequent labour market outcomes. In this paper, we study an important but under explored migration incentives arising from the matching outcomes in the marriage market and the gender differences in responding to such incentives. To achieve identification, we exploit the setup of special economic zones (SEZs) as a pull force and China's accession to the World Trade Organization (WTO) as a push force that exogenously trigger urbanization across locations, which leads to a unique feminization phenomenon during this process. The paper highlights important distributional implications on gender inequality and spatial disparity during the rapid urbanization process.

Table of Contents

Acknowledgement

First and foremost I would like to express my deepest gratitude to my supervisors Professor Zhenlin Yang and Professor Jing Li. They are always selfless in giving me research guidance with great patience. Also, their passion and professionalism in doing research make me clear about what kind of person I should expect myself to be. I'm so lucky to have them as my role models. This thesis would not have been possible without their encouragement and support. In addition, I sincerely thank my thesis committee, Professor Jungho Lee, and Professor Ying Deng for their invaluable comments.

I would like to thank our programme directors Professor Nicolas L. Jacquet and Professor Anthony Tay for their great leadership and support. Also, I would like to thank Ms. Qiu Ling Thor and Ms. Amelia Tan. They are always nice and helpful.

Last but not least, I am grateful to my family for their unconditional support. I would like to give special thanks to Shuhang Chen for the understanding, accompany, and love. Thank you.

Dedication

This thesis is dedicated to the people who supported and shaped me over the years. You made me who I am. Thank you.

Chapter 1

Identifying Knowledge Spillovers from Universities: Quasi-experimental Evidence from Urban China

1.1 Introduction

Economists and policy makers have long stressed the importance of higher education institutions in fostering economic growth (Acemoglu 1995; Redding 1996; Andersson et al. 2004, 2009; Aghion et al. 2009). The common belief is that universities not only train high-skill labor, but also disseminate knowledge and promote productivity in local communities (Valero and Van Reenen 2019; Andersson et al. 2009). Whereas previous research has shown the impacts of universities on the development of certain industries and local productivity, we have limited understanding on the causal role that universities play in facilitating knowledge-based externalities and on the geographic scope of such externalities (Kantor and Whalley 2014, 2019). The central challenges that have limited progress in the literature are the endogeneity concerns and the difficulty in distinguishing knowledge spillovers from other potential channels. This paper takes a geographic approach, combined with a quasi-experimental setting, to resolve the challenges and identify the role of knowledge spillovers from universities.

We examine the causal impact of university activity on the creation of local patents and new products by taking advantage of a unique quasi-experiment in China that has exogenously expanded higher education institutions since 1999. We exploit a structural break in the university-innovation relationship induced by the policy shock to uncover the localized nature and striking geographic attenuation of university spillovers at a very refined geographic level (within 2-3 km).[1] We achieve this goal by utilizing novel datasets that contain comprehensive information on patents and new products of firms geocoded at the address level. The uncovered geographic nature of the impact allows us to identify knowledge spillovers from universities by building on the general consensus that idea flows rely heavily on spatial proximity.[2] By further merging our core datasets with patent citation information, we also reveal direct evidence of knowledge outflows from universities and striking spatial decay patterns of the citation links. Our findings unanimously point to the importance of knowledge spillovers in fostering innovation in close proximity to education and research institutions.

The evidence on knowledge spillovers from universities helps researchers better understand the role of universities in the economic growth process. It is widely acknowledged that research and development (R&D) played a central role in advancing the world technology frontier and contributed to continued economic growth over the past 200 years (Acemoglu 2008). However, in innovation-based growth models, the R&D production function has been taken as a reduced-form representation and the specific steps leading to practical innovations is not yet clear.[3] Pre-

[1]While previous studies have documented the localized nature of university spillovers at the scope of cities or counties (Jaffe 1989; Audretsch and Feldman 1996; Anselin et al. 1997; Andersson et al. 2004, 2009; Kantor and Whalley 2014; Liu 2015; Kantor and Whalley 2019; Hausman 2022), none has studied the spillover effects at the refined geographic level as we undertake in this paper. The extension to this geographic level is important in identifying knowledge spillovers as one of the mechanisms that contribute to the impact of universities on local innovation.

[2]An extensive literature emphasizes that knowledge spillovers decay rapidly within narrowly defined geographic space (Jaffe et al. 1993; Rosenthal and Strange 2003, 2005, 2008; Arzaghi and Henderson 2008; Combes and Gobillon 2015; Li et al. 2022; Baum-Snow et al. 2021). This is because gains from exchanging knowledge and information rely heavily on close-range face-to-face contact. The geographic approach to identify the mechanisms of agglomeration externalities has been emphasized in Rosenthal and Strange (2020) and validated in Li et al. (2022).

[3]Externalities from human capital and innovation had a scientific revival with the endogenous growth models starting with Romer (1986, 1990), Lucas (1988), and Grossman and Helpman (1991). Jaffe (1986, 1989) modeled a simple production function using industry and university research as inputs. Both studies found significant and positive effects of university research on outputs.

sumably, research outputs from research institutions serve as a key first step leading to innovative ideas that are then converted into innovative products. By tracing the impact of university activities on patents, patent citations, and new products, we document the role of knowledge spillovers from universities in the innovation process. This process of transforming fundamental knowledge into patentable findings and practical products forms the cornerstone of the R&D production function that is at the center of innovation-based economic growth theory.

Understanding the presence and the spatial scope of university spillovers in promoting local innovation also has important policy implications. First, it justifies public investments on higher education that has witnessed enormous growth in recent decades (Schofer and Meyer 2005).[4] Second, the extent to which education investment spills over to benefit surrounding firms provides guidance for creating technology hubs near education institutions. Evidenced by the salient example of Cambridge's Kendall Square near Harvard University and the Massachusetts Institute of Technology, policy makers have formed a general consensus that proximity to universities is a key condition for a vibrant high-tech community. Yet, a careful policy design requires a good understanding on how quickly the positive externalities decay with geographic distance. If university spillovers decay slowly, social planners may not have to endure high congestion costs in close proximity to research institutions to exploit the spillover benefits. If, however, the positive externalities decay quickly, policy makers would need to carefully gauge policy parameters to balance the spillover benefits with rising congestion costs.

Empirically, it is challenging to identify the causal impact of university spillovers on local innovation activities. One possible endogeneity concern resides in the presence of persistent local unobserved amenities that attract both premier universities and productive firms. In addition, business activities also reversely impact nearby universities and academic research through collaborations with or donations to universities (Bils and Klenow 2000). We address the concerns by exploiting a unique national university expansion policy resulted from an unanticipated economic stimulus plan from the central government in China. The policy introduced an exoge-

[4]For example, in 2017, the Ministry of Education of China spent 1,110.9 billion yuan (about US$170.9 billion using the exchange rate in December 2017) on higher education (http://www.moe.gov.cn/).

nous structural break in city-specific university capacity that is presumably independent of local economic conditions. We make use of the kinked relationship created by the shock to identify the impact of the university expansion in a difference-in-differences framework, drawing on cross-sectional variations in the exposure to the shock determined by the university capacity prior to the shock.

More important, the core element of our empirical analysis is the focus on within-city variations to characterize the geographic nature of university spillovers and to identify the role of knowledge spillovers. This within-city focus allows us to adopt a triple-differences approach in which we control for a rich set of interacting fixed effects to tighten our identification. Specifically, to capture the spatial attenuating features at very refined geographical levels, we examine the impact of university expansion on surrounding industrial innovation activities within 0.5 km, between 0.5 km and 1 km, between 1 km and 1.5 km, and so on, extending up to 5 km or 10 km, depending on the specific model. We control for year by ring, year by city, and city by ring fixed effects to absorb unobserved local demand shocks or factors related to either China's World Trade Organization (WTO) accession or reduction in internal migration and trade costs.[5] Our focus on the localized geographic nature of the impact allows us to shed light on the geographic scope and the underlying mechanism of university spillovers. As existing studies have shown how fast knowledge spillovers decay over space, taking the analysis to this level of geography is essential.

In the empirical analysis to follow, we document the extent to which proximity to academic universities in China affects nearby patent generation and cross-patent citations. We utilize detailed patent-level data between 1995 and 2007 from the National Intellectual Property Administration of China and patent citation links scraped from Google Patents to achieve this focus. Patenting is one of the best proxies for innovation and is widely used to capture knowledge creations. Since Jaffe et al. (1993), the literature has taken advantage of patent citation links to trace the paper trails of knowledge flows.[6] We rely on citation links to highlight direct

[5]We address further concerns on the possible presence of city- and location(distance)-specific unobserved time-varying factors by taking advantage of information on nearby patents that cite university patents. We elaborate on this point in Section 3.

[6]Although the case-control approach in Jaffe et al. (1993) faces challenges and is refined in several follow-up efforts, the approach of following patent citations to trace knowledge flows is

knowledge flows from universities to nearby industrial firms. Moreover, a comparison between nearby patents that cite university patents and those that cite patents far away from universities helps address concerns on the possible presence of city- and location (distance)-specific unobserved time-varying factors that may contaminate our triple-differences identification.

Despite the benefit of detailed information on patents, patenting represents an intermediate step rather than the final economic output in the process of converting new ideas to new products. To mitigate this concern, we take advantage of previously under-explored information on new firm product sales reported in the Annual Survey of Industrial Firms (ASIF). According to ASIF, "products included in the category of new product sales are those that are new in relation to the reporting firm's prior product mix." Hence, new product sales in ASIF better reflect the ultimate outcome of the innovation process: the commercialization of technical ideas. Other firm-level surveys rarely capture this information on new product.[7] It provides a unique opportunity to examine the impact of university expansion on a direct measure of downstream outputs produced using knowledge and ideas.

We obtain the following results. First, university innovation activities increase nearby patents, and the impact decays sharply with geographic distance. In particular, we find that the level of patenting activities reduces by about 80 percent when moving from within 0.5 km to 0.5-1 km of a university. The impact reduces by another 65 percent when moving from 0.5-1 km to 1-1.5 km of a university. The sharp decline stops roughly at 2 km away, and the attenuation slope flattens out thereafter. Second, we find that the spatial attenuation of university spillovers is ubiquitously present in different regions and industries in China but is more pronounced in the Eastern region and for industries more reliant on high-skilled labor. Third, we find that university expansion increases nearby industry patents that cite university patents. The knowledge outflows captured by citation links also decay quickly across space and stabilize beyond the 2 km radius. The spatial attenuation pattern, however, is not present for the number of times when nearby industry

widely recognized (Thompson and Fox-Kean 2005; Thompson 2006; Murata et al. 2014; Figueiredo et al. 2015).

[7] A few studies use new product announcement data from the U.S. Small Business Administration to examine innovation (Acs and Audretsch 1988; Acs et al. 1994; Feldman and Audretsch 1999; Acs et al. 2002). That data, however, are only available for 1982 and are also limited in scope.

patents cite patents far away from universities. Last, further analysis suggests that university expansion boosts new products from firms and that the impact follows a similar spatial decay. This effect is more pronounced for high-skilled intensive industries and private firms than for low-skilled intensive industries and state firms.

Our study contributes to two sets of literature. First, this paper joins the literature on knowledge spillovers and agglomeration economies. Since Marshall (1890), researchers have attributed the micro-foundations of agglomeration externalities to the sharing of goods, people, and ideas—otherwise labeled as intermediate input sharing, labor market pooling, and knowledge spillovers (Duranton and Puga 2004; Holmes 1999; Glaeser and Maré 2001; Moretti 2004; Ellison et al. 2010). It is also recognized that different microfoundations are associated with different spatial attenuation of agglomeration externalities (Rosenthal and Strange 2004; Combes and Gobillon 2015; Li et al. 2022).[8] Rosenthal and Strange (2020), in particular, emphasizes the convenience of identifying the nature of agglomeration externalities by relying on the observed attenuation patterns. Baum-Snow et al. (2021) interprets the micro-geographic level rapid spatial decay in productivity spillovers as explained by learning or knowledge transfer. Hence, the fast attenuation speed documented in our paper points to the important role of knowledge spillovers in university spillover benefits.

Second, we contribute to the literature on the impact of research institutions and academic research on local economic outcomes. Previous studies have focused on a range of economic outcomes in the context of developed countries. For instance, Jaffe (1989) and Anselin et al. (1997) examine the effects of university research on local innovations in the United States. Andersson et al. (2004, 2009) investigate the impact of educational investment on productivity and innovation in Sweden. Kantor and Whalley (2019) uses historical establishment of agricultural experiment stations in the United States to evaluate the impact of proximity to research on agricultural productivity.[9] However, the geographic unit of analysis is mostly at the city, county,

[8]For instance, industries that rely heavily on knowledge spillovers as the main agglomeration force often require close-range face-to-face contact, which implies a rapid spatial decay of agglomeration spillovers; industries that cluster mainly because of input-output linkages could have agglomeration externalities decay slowly and extend to a larger spatial scale.

[9]Other studies in this strand of literature include Aghion et al. (2009), Kantor and Whalley (2014), Liu (2015), Andrews (2019), and Hausman (2022).

municipality, state, or region level, which prohibits researchers from understanding the micro-geographic scope of university spillovers and drawing conclusions on the channel through which the localized spillovers take place. We extend the literature by focusing on a developing country context and documenting a sharp spatial attenuation of the impact of university expansion on intermediate innovation outcomes (patents and citation links) and final output measures (new products).

The rest of the paper is organized as follows. Section 2 introduces the institutional background of the university expansion in China. Section 3 lays out the empirical framework and identification strategies. Section 4 describes data and variables. Section 5 presents the empirical results on patents. Section 6 presents the results on patent citations and new products. We conclude in Section 7.

1.2 Institutional Background

In this section, we introduce China's higher education system and discuss the policy background of the university expansion that started in 1999.

China's higher education system is under central planning since its establishment in the 1950s. The Ministry of Education (MOE) in the central government is the sole entity that makes admission plans for all universities based on the national economic development plan. High school graduates are admitted to different universities based on their performance on a unified national college entrance examination. This central planning feature governs that the implementation of higher education policies follows a top-down approach, and the intensity of the policy is usually not responsive to economic environment at the local level.[10] The radical university expansion that started in 1999 is one such example and was unanticipated at the time.

Before 1999, the development of China's higher education institutions was steady and smooth.[11] However, the onset of the 1997 Asian financial crisis and the massive

[10] China's higher education system is different from the systems in many Western countries, such as the United States. or example, almost all prestigious universities in China are public universities, whereas many prestigious universities in the United States are private. In addition, the financial support for higher education is almost entirely provided by the MOE in China, whereas fundraising plays a significant role in financing the universities in the United States.

[11] China's higher education system went back to normal after the Cultural Revolution ended in 1976. In the 1990s, the Ministry of Education guided China's higher education sector under a theme called "steady development." The number of enrolled students increased with an average growth

layoffs resulted from the state-owned enterprise (SOE) reforms in the late 1990s raised concerns about a recession and triggered a need to expand the higher education sector to stimulate the domestic demand for educational services and other related consumption.[12] University expansion was also believed to postpone the entry of high school graduates into the labor market, which may otherwise exacerbate the already-high unemployment rate (Che and Zhang 2018).

In June 1999, the MOE and the National Development and Planning Commission jointly announced a new higher education recruitment plan, with expected new students of 1.53 million in 1999—a 42 percent year-on-year increase. In the meantime, college tuition fees increased by 15-20 percent across different regions. The revenue from tuition became an important financial resource for universities. The central government also shifted more resources to higher education to accommodate the huge increase in university scale. From 1998 to 2000, the science and technology funding and national expenditure on higher education increased from 8.2 billion yuan to 14.3 billion yuan and from 33.6 billion yuan to 49.1 billion yuan, respectively. From 1998 to 2000, the number of university teachers also rose by 55,519, more than fourfold of the increase from 1990 to 1998.

The expansion was unanticipated by the general public and local governments. The new plan would impose huge impacts on the college entrance examination in July (one month later) and the new academic semester starting in September (three months later). The time left for the government to distribute the enrollment quota was pressing. Official documents suggest that the quota allocation across cities mainly depended on the national expansion plan and existing universities' physical and logistical capacity at the city level. The quota allocation rules also present strong inertia as the radical expansion continued in the following years. Therefore, the expansion led to an exogenous structural break in a city's higher education scale,

rate of 7 percent between 1977 and 1998. From 1990 to 1998, the number of university students increased from 2.06 million to 3.41 million, and the number of university teachers rose slightly from 0.395 million to 0.407 million.

[12]Min Tang, a famous economist in the Development Research Center of the Asian Development Bank, originally proposed the university expansion policy. In November 1998, Mr. Tang, along with his wife Xiaolei Zuo, wrote an open letter to the central government, in which they appealed for doubling the higher education enrollment in three years. They also suggested that China stop offering free higher education and require students to pay tuition fees. They believed that those actions would help generate demands in relevant economic sectors and stimulate the nation's economy. The letter can be viewed at http://finance.sina.com.cn/review/20041023/15201102716.shtml.

and the magnitude of the structure break depended on the city's university resources prior to the shock.

Figure 1.1 depicts various aspects of the structural break in China's higher education sector induced by the policy shock. Panels A-D present the numbers of university teachers, university students, university entrants, and university graduates from 1990 to 2010. Before 1999, the growth rate of those numbers was low and steady. However, a clear trend break exists in the time series of the numbers of university teachers, university students, and university entrants in 1999 and the number of university graduates in 2003. The scale of universities in China increased dramatically after 1999. In particular, the number of university teachers in 2010 was more than three times of the number in 1998. In Panel E, the national higher education expenditure increased more than threefold from 1998 to 2006. Panel F shows that the science and technology funding for the higher education sector increased by a factor of over 11 from 1998 to 2010. The expansion dramatically increased the research resources to universities, at both the aggregate and per teacher level.

In Figure 1.2, we plot the correlation between the extent of university expansion from 1999 to 2007 and the scale of higher education before the expansion at the city level. Specifically, in Panel A, we plot the increase in the number of university teachers between 1999 and 2007 in each city against the number of university teachers in each city in 1990. We do the same for the number of university students in Panel B. There is a clear positive correlation between the expansion in university scale and the pre-existing university students and teachers before the expansion at the city level. The pattern confirms that, during the expansion period, the enrollment quota was allocated to different cities mainly based on the city-level pre-existing physical and logistical capacity of the higher education sector. The increase in enrollment quota further induces universities to gain more funding, upscale the teachers, and eventually expand the research capacity.

In sum, the higher education expansion policy followed a top-down approach and created a positive exogenous shock to university scale. The extent of the expansion in each city was largely determined by the national expansion plan and existing universities' capacity before the expansion. Several studies find support for the exogeneity of this national policy to the local economic environment (Che and

Zhang 2018; Li et al. 2017; Rong and Wu 2020).[13] We provide similar evidence in Appendix Figure A1 that the extent of university expansion in a city is not predicted by the growth of patents, GDP and firm TFP in the city before the expansion. We use this policy shock to form a difference-in-differences and a triple-differences research design.

1.3 Empirical Framework

We face three empirical challenges. The first challenge pertains to the identification of the impact of university activity on local economic outcomes. An endogeneity concern arises from that university activity does not occur randomly. For instance, there may exist location-specific unobserved characteristics that attract innovative firms and research-oriented universities simultaneously. Alternatively, the nearby presence of innovative firms may reversely affect the activities of universities through knowledge spillovers from industrial firms to universities, through donations to universities and collaborations, or through increased local demand for a university-trained labor force.

We address the endogeneity concern by utilizing the university expansion policy in China as a quasi-experiment. As explained in Section 2, the policy created an unanticipated structural break in the intertemporal development of universities, which allows us to identify the causal impact of the university expansion in a difference-in-differences framework. We examine the extent to which the policy shock induces an expansion in university patenting and the extent to which it spills over to affect citywide industrial patenting activities. The regression equation is specified as follows:

$$Outcome_{c,t} = \beta \times (Treatment_c \times Post_t) + \alpha_c + \gamma_t + \varepsilon_{c,t}, \tag{1.1}$$

where $Outcome_{c,t}$ represents $UniversityScale_{c,t}$, the numbers of university teachers, university students, or university patents in city c and year t, or $IndustryInno_{c,t}$, the number of collaboration patents or industry patents in city c and year t. $Treatment_c$

[13] Che and Zhang (2018) shows that the annual growth rates of gross domestic product and annual admission are uncorrelated at the provincial level for the period of 1995-2011. They also show that the correlations between the growth of new college graduates in 2001-2003 and the growth of provincial GDP and firm TFP are small and statistically insignificant.

is the number of university teachers (or students) in city c in year 1990, a proxy for treatment intensity. $Post_t$ is a dummy variable that equals 1 if year t is 2000 or after. α_c and γ_t are city and year fixed effects, and $\varepsilon_{c,t}$ is the error term.

This identification strategy draws on cross-sectional variation in the exposure to the shock determined by the university capacity prior to the shock. The identification assumption is that the evolution of outcome variables in cities with larger expansions should not vary systematically from cities with smaller expansions in the absence of the expansion, conditional on included control variables. In other words, any pre-existing trends should be properly controlled for. We discuss the validity of the identification assumption in more detail below. Also note that if we are willing to make additional assumption that the expansion impacts industry innovation only through increasing university scale, we can consider the impact of the expansion on $UniversityScale_{c,t}$ as the first-stage effect, and the impact on $IndustryInno_{c,t}$ as the reduced-form effect, in a standard Wald difference-in-differences setup (Duflo 2001; Bhuller et al. 2013).[14]

The second empirical challenge hinges on the core of the paper—identifying the role of knowledge spillovers as an important channel contributing to the impact of universities. That is, even if the causal impact of university activity on local economy is convincingly justified, it is not clear whether the impact is channeled through knowledge spillovers. For example, an increase in university scale may accompany an increase in the supply of college graduates if graduates prefer to work in the area where they attend college (Card 1995). Increased high-skilled labor could improve local economic outcomes directly (Che and Zhang 2018). Hence, the challenge is how to safely disentangle the role of knowledge spillovers from other mechanisms.

We tackle this challenge by focusing on the extremely localized effects of universities. Previous studies have shown that knowledge spillovers tend to decay rapidly across space, while other benefits of agglomeration, such as labor market pooling, operate at a much larger geographic scope (Rosenthal and Strange 2003,

[14]The effect of university innovation capacity on industrial innovation activities can be retrieved by taking the ratio of the reduced-form and the first-stage estimates or by two-stage-least-squares (2SLS) estimation with direct inference. Because the assumption of exclusion restriction is harder to justify, we focus on the difference-in-differences model, but we also report the 2SLS estimates in the appendix tables.

2005; Arzaghi and Henderson 2008; Li et al. 2022). In particular, Arzaghi and Henderson (2008) depict sharply attenuating knowledge spillovers and networking benefits that deplete at 750 meters away. As noted in Carlino and Kerr (2015), this spatial approach "represents an important precedent for future research related to innovation more directly." Indeed, the focus on the geographic nature of university spillovers is convenient to disentangle the important role of knowledge spillovers. It is difficult to imagine that alternative channels, such as the labor market channel, would dissipate dramatically at a short distance away from a university.[15] In the online appendix, we present a simple conceptual framework to formalize the identification of knowledge spillovers by drawing on the localized nature of knowledge spillovers.

We achieve this geographic focus econometrically by specifying a rich set of concentric ring variables that capture innovation activities at various distances from research universities. Each concentric ring spans 500 meters. We include 10 or 20 rings to cover places up to 5 km or 10 km away from a university, depending on the specific model.[16] This additional source of within-city variation allows us to identify the spatial attenuation of university spillovers in a triple-differences framework, specified as follows:

$$IndustryInno_{c,r,t} = \sum_{r=1}^{9} \beta_r \times (Treatment_c \times Post_t \times Ring_r) + d_{c,r} + d_{c,t} + d_{r,t} + \varepsilon_{c,r,t}, \tag{1.2}$$

where $IndustryInno_{c,r,t}$ represents the number of industry patents in city c, ring r, and year t; $Treatment_c$ and $Post_t$ are defined the same as before; $Ring_r$ is a dummy variable that equals 1 if the patents are in the concentric ring r and 0 otherwise (ring 10 is set as the reference group and is omitted); $d_{c,r}$, $d_{c,t}$, and $d_{r,t}$ are city by ring, city by year, and ring by year fixed effects, respectively.

The ability to include all interactive fixed effects is crucial for identifying the geographic nature of university spillovers. China experienced dramatic economic

[15] It is possible that highly innovative firms are attracted to the close proximity of research universities to draw on the spillover benefits. As a result, firms closer to universities may disproportionately hire university graduates. However, we should be careful and not interpret the increased innovation activities near universities as a mere consequence of disproportionately allocated high-skilled labor since the latter is an equilibrium outcome of knowledge spillovers and serves as a channel through which knowledge spillovers benefit nearby innovation in a self-reinforcing process.

[16] Section 4.2 provides a detailed explanation on the construction of the rings.

reforms in the past few decades. For instance, since the early 2000s, the Chinese government has undertaken policy reforms and infrastructure investments that have substantially reduced the costs of internal migration and trade. China also joined the WTO at the end of 2001, which led to large reductions in international trade costs. Those reforms contributed to a significant growth in aggregate productivity and may drive increases in innovation activities (Brandt et al. 2017; Tombe and Zhu 2019). We address those potential confounding factors by including city by year, ring by year, and city by ring fixed effects in a generalized triple-differences framework. In particular, city-level time-varying unobservables, the main confounding factor in the difference-in-differences model, are controlled for by city by year fixed effects. The remaining unobserved factors conditional on those demanding interacting fixed effects are unlikely to systematically impact innovation activities at different distances from universities. Thus, the triple-differences strategy relies on weaker identification assumption than the difference-in-differences strategy.

The third empirical challenge is the measurement of innovation. Conceptually, innovation should comprise generation of new ideas and conversion of ideas into commercial products. The new ideas generated could sometimes result in patents. Therefore, it is natural to use patent as a proxy for innovation. Plus, patent data are also publicly available and contain rich details. However, two potential concerns exist: (1) patents do not directly reflect knowledge flows and (2) they are an intermediate step in the innovation process and do not capture the ultimate economic value of the invention (Acs et al. 2002).[17] Because patents and new products do not necessarily collocate, we need to interpret the patent-based evidence with caution (Feldman and Kogler 2010).

To mitigate measurement concerns associated with using patent counts as proxies for innovation, we supplement our patent analysis with subsidiary analyses on two additional measures: patent citation links and new commercial products. We examine the incidences when industry patents cite university patents as direct evidence of knowledge transfers from universities. We also take advantage of previously under-explored information on firms' new commercial products to reveal the

[17] Based on Acs and Audretsch (1988), Griliches (1979) and Pakes and Griliches (1980), "patents are a flawed measure (of innovative output) particularly since not all new innovations are patented and since patents differ greatly in their economic impact."

impact of university spillovers on the final products.

We estimate the following triple-differences specification to capture the spatial decay of patent citations:

$$Cite_{c,r,t} = \sum_{r=1}^{9} \beta_r \times (Treatment_c \times Post_t \times Ring_r) + d_{c,r} + d_{c,t} + d_{r,t} + \varepsilon_{c,r,t}, \tag{1.3}$$

where $Cite_{c,r,t}$ represents the number of cases when industry patents in city c, ring r, and year t cite university patents. The rest variables are defined in the same way as in Equation (1.2).

To explore the impact of university expansion on nearby new products at the firm level, we estimate the following specification:

$$\begin{aligned} NewProduct_{i,c,r,t} = &\sum_{r=1}^{9} \beta_r \times (Treatment_c \times Post_t \times Ring_r) \\ &+ d_{c,r} + d_{c,t} + d_{r,t} + \boldsymbol{X}_{i,c,r,t}\boldsymbol{\rho} + \varepsilon_{i,c,r,t}, \end{aligned} \tag{1.4}$$

where $NewProduct_{i,c,r,t}$ represents the new commercial product ratio of firm i in city c, ring r, and year t; $\boldsymbol{X}_{i,c,r,t}$ is a set of firm-specific controls, including the age of a firm, fixed assets, a dummy for whether a firm is an SOE, and the employment size; and $\varepsilon_{i,c,r,t}$ is a firm-specific error term. We define the rest variables in the same way as in Equation (1.2).

The set of differencing strategies stated above relies on the identifying assumption that, conditional on included fixed effects and other controls, the evolution of the outcome variables in cities with larger expansions should not vary systematically from cities with smaller expansions in the absence of the expansion (difference-in-differences setup), and, in the event that there exist systematic variations across cities, such counterfactual differences do not vary across rings (triple-differences setup). A natural check on the validity of such assumption is whether the pre-trends are parallel. We perform event-study analyses to check on this assumption and also to capture the dynamics of the treatment effects. The event-study specification for the city-level analysis is as follows:

$$\begin{aligned} Outcome_{c,t} = &\sum_{t=1995}^{1998} \beta_t \times (Treatment_c \times Year_t) \\ &+ \sum_{t=2000}^{2007} \beta_t \times (Treatment_c \times Year_t) + \alpha_c + \gamma_t + \varepsilon_{c,t}, \end{aligned} \tag{1.5}$$

where $Outcome_{c,t}$ represents $UniversityScale_{c,t}$ or $IndustryInno_{c,t}$; $Year_t$ is a set of year dummies that equals 1 if year equals t and 0 otherwise. Year 1999 is set as the base year and is omitted. We define the rest variables the same as in Equations (1.1).

We also estimate event-study model for the ring-level analysis as follows:

$$\begin{aligned} IndustryInno_{c,r,t} = & \sum_{r=1}^{9} \sum_{t=1995}^{1998} \beta_{r,t} \times (Treatment_c \times Year_t \times Ring_r) \\ & + \sum_{r=1}^{9} \sum_{t=2000}^{2007} \beta_{r,t} \times (Treatment_c \times Year_t \times Ring_r) + d_{c,r} + d_{c,t} + d_{r,t} + \varepsilon_{c,r,t}, \end{aligned} \tag{1.6}$$

where $Year_t$ is a set of year dummies that equals 1 if year equals t and 0 otherwise. We define the rest variables in the same way as in Equation (1.2).

However, as discussed in detail in Section 5, the event study analysis reveals that the pre-treatment trends seem to be not sufficiently controlled for by the included fixed effects and controls in both the city-level and ring-level analyses. This could be explained by that cities experiencing more intensive university expansions may also have been adopting more innovation-promoting and growth-enhancing policies before the expansion and such efforts could be more directed towards areas close to existing innovations than far-away areas.

We undertake a collection of efforts to address this concern. First, we examine whether the university expansion induces a slope change in variables of interest by estimating a trend break model, following Almond et al. (2019). As the policy created an unanticipated structural break in the intertemporal development of universities and the magnitude of the structural break is independent of unobserved local economic conditions, we rely on a kinked relationship to identify the impact of universities on local innovation through the channel of knowledge spillovers. The model at the city level is specified as follows.

$$\begin{aligned} IndustryInno_{c,t} = & \beta \times (Treatment_c \times Trend_t) \\ & + \gamma \times (Treatment_c \times Trend_t \times Post_t) + \alpha_c + \gamma_t + \varepsilon_{c,t}, \end{aligned} \tag{1.7}$$

where $Trend_t$ is a trend variable defined as the patent application year minus 1999; The rest variables are defined the same as in Equation (1.1). The coefficient β measures the difference in trends associated with cities of different treatment intensity prior to the university expansion. The coefficient γ measures the post-expansion slope change in the outcome variable relative to the pre-expansion trend.

The trend break model at the ring level is specified as follows.

$$\begin{aligned} IndustryInno_{c,r,t} = & \sum_{r=1}^{9} \beta_r \times (Treatment_c \times Trend_t \times Ring_r) \\ & + \sum_{r=1}^{9} \gamma_r \times (Treatment_c \times Trend_t \times Ring_r \times Post_t) + d_{c,r} + d_{c,t} + d_{r,t} + \varepsilon_{c,r,t}, \end{aligned} \tag{1.8}$$

The variables are defined the same as before. The coefficient β_r measures the difference in trends for ring r associated with cities of different treatment intensity prior to the university expansion. The coefficient γ_r measures the post-expansion slope change in the outcome variable relative to the pre-expansion trend for ring r.

Second, we strip away the city-specific or city-ring-specific pre-expansion linear time trend as a control strategy before we run difference-in-differences and triple-differences specifications, following the approach in Bhuller et al. (2013), Monras (2019) and Garcia-López et al. (2020).[18] Specifically, we estimate a city-specific or city-ring-specific linear trend using the pre-expansion sample (namely, 1995-1999) for our city-level and ring-level regressions, respectively. We then extrapolate pre-expansion time trends to the post-expansion sample and subtract out the estimated linear trend from the observations after treatment. We use the trend-free outcome measures as the dependent variables in the city-level and the ring-level analyses. We re-estimate a trend-free event study model to verify that the residualized pre-trends are parallel and also to depict the intertemporal dynamics of the impact. The event study model further tightens our identification by leveraging on the sharp timing of the university expansion and high frequency measurement of the outcomes.

Third, we conduct a set of robustness checks to corroborate our main results. In the first robustness check, we follow Dobkin et al. (2018)'s parametric event study approach to augment our baseline city-level and ring-level specifications with city-specific and city-ring-specific linear trends, respectively. This approach is conceptually the same as subtracting out the estimated linear trend elaborated above (Goodman-Bacon 2018, 2021; Rambachan and Roth 2022). In the second robustness check, we follow Rambachan and Roth (2022) to obtain robust inference after specifying how different the post-treatment violations of parallel trends can be from

[18] As stated in Monras (2019), this is a valid identification strategy if in the absence of the treatment the outcome variables would have evolved following the linear trend implied by the periods preceding the treatment event.

the pre-treatment differences in trends. This approach also allows us to conduct sensitivity analyses showing whether a causal conclusion can be drawn under various restrictions on possible violations of the parallel trend assumption. Details are discussed in Section 5.

Last, we further tighten our identification by drawing on variations in different types of citation flows. A possible argument against the identification of knowledge spillovers even with our most sophisticated generalized triple-differences model is that there may exist unobserved city- and location (distance)-specific time-varying factors that are correlated with the increase in location-specific innovation activities after the university expansion. Such a hypothetical scenario is possible but very unlikely given the rare coincidence of multiple co-evolving factors after controlling for a demanding set of fixed effects. Despite so, we address this concern by drawing on the information on patent citation links to show that the spatial pattern persists only for citation links of industry patents citing university patents but not for citation links of industry patents citing patents far away from universities. Otherwise, if unobserved co-evolving factors drive the spatial pattern of overall patenting activities and citation behaviors, we would observe similar patterns for both types.

1.4 Data, Variables, and Summary Statistics

1.4.1 Data

We use four primary datasets. The first dataset is a patent database obtained from the National Intellectual Property Administration of China (CNIPA). This dataset covers a complete list of patents granted between 1995 and 2007 in China. The data provide detailed information for each patent, such as inventor's name and affiliation, address of the patent, application date, approval date, patent ID, International Patent Classification (IPC) number, and patent type. There are three types of patents in the database: invention patent, utility model patent, and design patent.[19] We focus on invention patents because they represent the most innovative type. Overall, there were 553,248 invention patents granted in China between 1995 and 2007. We use

[19]Invention patents require inventive technological improvements or new uses. Thus, invention patents have the highest standard of novelty. The other two types of patents are related more to the structure (utility model patent), shape (utility model and design patent), and design (design patent) of an object and have fewer requirements for inventiveness.

invention patents to measure innovation activities inside and outside of universities.

The second dataset is extracted and compiled from four different statistical yearbooks of China. The first source is the China City Statistical Yearbook between 1996 and 2008 from the National Bureau of Statistics (NBS) of China. This collection provides information on various prefecture-city-level attributes by year, such as the number of university teachers and students.[20] The second source is the Educational Statistics Yearbook of China. We obtain the number of university entrants and graduates at the provincial and national level for each year from this yearbook. The third source is the Educational Finance Statistical Yearbook of China, which reports the higher education expenditures from 1995 to 2006. The fourth source is the Compilation of Statistical Data on University Science and Technology Resource, which provides information on the science and technology funding for higher education from 1991 to 2010. We use the number of university teachers and students from the first source as proxies for university scale or research capacity. The other three sources help us summarize aggregate trends for various aspects of the university expansion.

The third dataset is a patent citation database that is scraped from Google Patents. Google Patents is a search engine from Google that indexes patents and patent applications from all around the world. We searched for all patents granted in China. For each patent, we collect its basic information and patent citations. Then, we match the data to our patent database from CNIPA. This gives us a patent citation matrix about whether a patent cites another patent. We treat the patent citation links as the paper trail of knowledge flows and use them to identify knowledge spillovers from universities.

The final dataset is the ASIF of China from 1998 to 2007. This dataset is also from the NBS of China. The ASIF is an annual panel that covers all SOEs and the non-SOEs with annual sales exceeding 5 million yuan.[21] The data provide detailed firm-level attributes, including firm name, firm address, legal unit code, legal representative name, industry classification, opening year, ownership type, fixed capital,

[20]The statistical yearbooks report the statistics for the previous year.

[21]The ASIF contains many missing values after 2007. In addition, starting in 2011, the sampling cut-off increased to 20 million yuan of annual sales, which changes the sample composition and makes comparisons across years challenging.

output value, and employment size, among others.[22] A unique advantage of the ASIF is the exact firm addresses provided in the data. We geocode the addresses and pin the firms into the concentric rings that we create.

We use the previously under-explored information on firm new products in the ASIF to measure the final commercialized outputs with new knowledge and ideas as inputs. The NBS defines a new product as a product that is produced for the first time at least within a province (Lu and Tao 2009). Based on an email correspondence with an officer at the NBS, "products included in the category of new product sales are those that are new in relation to the reporting firm's prior product mix. Products that involve the use of new principles, incorporate design improvements, utilize new materials, or embody new techniques constitute new products; existing products that are used for new functions or expand capabilities (e.g., production or speed) also constitute new products. Changes in a product's shape or minor changes in functionality do not constitute new products" (Jefferson et al. 2003). Other firm-level surveys rarely capture this measure of new product. It provides a unique opportunity to study final outputs from innovation. We use a firm's new product ratio as a proxy for innovation output and define it as the ratio of the dollar value of new products to the dollar value of total outputs.

1.4.2 Variables and Summary Statistics

In this section, we describe how we prepare our data for the empirical analysis and present the basic summary statistics. For the city-level analysis, we create a city by year panel by matching patent counts at the city level to city-level attributes from the statistical yearbooks. The year of patenting refers to the year when the patent application is filed, as opposed to the year when the patent is granted. Our goal is to trace how the flow of knowledge impacts the creation of new ideas, and the application year is closer to the timing of new knowledge creation (Moretti 2021). We have 184 cities in the panel after removing observations with missing information.

Table 1.1 presents the summary statistics for the numbers of university teachers, university students, and the total and sub-classifications of patents at the city level in

[22] In the empirical analysis, we adjust all dollar variables using the national Consumer Price Index (CPI) so that they are comparable across years.

each year. Columns (2) and (3) show the average number of university teachers and students. The city-average growth trends are similar to the national trends in Figure 1, showing a dramatic boom in university scale after 1999. Column (4) reports the average number of patents at the city level in each year. The number of patents also increased dramatically from 2000, which matches the timing of the university expansion. We further decompose patents into three mutually exclusive categories. Column (5) shows the average number of university patents, which we define as patents filed solely by inventors affiliated with universities. Column (6) reports the average number of collaborative patents between universities and the private sector. Column (7) reports the average number of patents that are filed solely by inventors from non-university entities. The three types of patents all experienced a sharp increase from 2000.

For the ring-level analysis, we create a panel at the city-year-ring level. To construct the rings, we compile a list of university locations as the centers of the rings in three steps. First, we manually search the locations for an exhaustive list of universities that are classified as "*Yiben*" universities in each city.[23] Second, we supplement the list with the locations of institutions in CNIPA that are classified as universities during our sample period.[24] Third, we add to the list the locations of other institutions or companies that have ever filed a joint patent application with a university during our sample period. This third step allows us to include possible university spin-offs in the university locations and avoids treating industry-university partnerships as spillovers.[25]

Then, we define the rings as a set of concentric rings around the universities locations. Specifically, we define one concentric ring for every 500 meters away from the center locations and have the rings extend up to 5 km or 10 km, depending on the specific model.[26] To identify the innovation activities within each concentric

[23]In China, universities are classified into several tiers. The tier of a university determines whether the university has priority when recruiting students. In general, "*Yiben*" (first tier) universities have the highest priority when recruiting students. "*Yiben*" universities also conduct the majority of research because they have better research and teaching capacity.

[24]This procedure may lead to multiple locations within the same university as the address filed in a patent application points to the exact building.

[25]Hall et al. (2003) documents industry-university research partnerships and suggests that the involvement of universities in industrial innovation benefits the outcome. However, we recognize that the patents generated from such partnerships should not be interpreted as spillovers from universities.

[26]We include 10 rings which extend up to 5 km in our baseline specifications. We include 20 rings to cover a broader geographic scope for robustness checks.

ring, we geocode the locations of patents and companies in CNIPA and ASIFs, and pinpoint them to the corresponding ring area. The patents generated at the center locations are not included in any rings. In the baseline ring-level analysis, variable $Inno_{c,r,t}$ is the number of patents in city c, ring r, and year t, where ring r refers to the concentric ring between the buffer zones $r-1$ and r.

We provide a graphic illustration in Appendix Figure A2 to show how we define the rings. In this example, the two locations at C belong to university I. The three locations at D and E belong to university II. Point A stands for a non-university entity that has direct collaboration with university I. Point B stands for a non-university entity that has direct collaboration with university II. We treat all these locations as the centers of a set of concentric rings. Each concentric ring spans a distance of 500 meters. The concentric rings, hence, are the outer envelopes that trace the rings of the same distance away from the center locations.

Table 1.2 presents the summary statistics for the number of patents within different concentric rings in each year across cities. First, we notice that the magnitude of the patent counts in the closest ring dominates that of the outer rings. For all years, the number of patents decays sharply as the distance to universities increases. This suggests that the overall innovation activities around universities are more intense than other areas. Second, a positive trend exists for all rings over time with a sharper increase after 2000. For example, the average growth rate of patent counts in ring 1 was 21.5 percent from 1995 to 1999, but it increased to 54.6 percent from 2000 to 2007. We also found similar but more muted patterns for outer rings.

1.5 Results on New Patents

1.5.1 City-Level Analysis on Patent Growth

We first examine the impact of universities on citywide innovation activities. In Table 1.3, we report the results from estimating Equation (1.1) when we use the number of university teachers in 1990 as the proxy for treatment intensity. The corresponding results when the number of university students in 1990 is used as treatment intensity are reported in Appendix Table A1.A. Column (1) of Table 1.3 suggests that cities with 1,000 more university teachers in 1990 experienced an additional increase of 341 university teachers after the university expansion. Columns (2) suggests that cities with larger university capacity in 1990 also experienced a larger increase in the total number of patents after the expansion.[27]

Next, we decompose the citywide total patent counts into the numbers of patents filed solely by inventors affiliated with universities, patents jointly filed by inventors affiliated with universities and inventors from industrial firms, and patents filed solely by industrial firms. Column (3) of Table 1.3 shows that cities with 1,000 more university teachers in 1990 experienced an additional increase of 33 university patents on average after the expansion. This suggests the expansion indeed boosted university innovation capacity, as represented by the number of university patents. Columns (4) and (5) show that cities with larger university capacity in 1990 also experienced a larger increase in collaborative patents and industry patents after the expansion. The impact on collaborative patents represents an important form of universities' contribution to the local economy by collaborating with other sectors.[28] The impact on industry patents implies potential spillovers from universities.

As discussed in the empirical framework, we can form a structural interpretation of the estimated coefficients in a Wald difference-in-differences setup, with additional assumptions. Specifically, dividing the reduced-form effect by the first-stage effect produces the Wald estimator of the impact of university's research capacity on industry patents. For example, in Table 1.3, the impact of university teachers on industry patents at the city level is 0.30 (101.75/341.02). Alternatively, the impact of university patents on industry patents at the city level is 3.05 (101.75/33.32). The magnitude of the effects is economically important: adding 100 more university patents to the average prefecture city increases the industry patents in the city by 305.[29]

To check on the parallel trend assumption and also to depict the dynamics of the treatment effects, we estimate event study models as in Equations (1.5). Appendix Figures A3 and A4 present the estimation results. In Panel (a) of Appendix Figure

[27] Additional results are presented in Appendix Tables A1.B-A1.C to show robustness when we add city-level control variables, such as the non-agricultural population, the proportion of employment in the manufacturing industries, and the proportion of employment in the service industries.

[28] As illustrated in Hall et al. (2003), research projects with university involvement tend to be in areas involving new science. The social benefits from the collaborated patents can be large.

[29] We present the two-stage least squares estimation results in Appendix Tables A4.A and A4.B with corresponding statistical inference.

A3, we show the dynamic effects of the university expansion on the numbers of university teachers and university patents using the number of university teachers in 1990 as the treatment intensity. Panel (b) shows the corresponding estimates using the number of university students in 1990 as the treatment intensity. Two patterns emerge. First, the university scale measured by the numbers of university teachers and patents do not present significant responses to variation in treatment intensity before the expansion when the number of university teachers in 1990 is used as the treatment intensity. However, there seems to be a small upward trend in university scale leading the expansion when the number of university students in 1990 is used as the treatment intensity. Second, the increases in the numbers of university teachers, university students, and university patents after 1999 are positively affected by the number of university teachers or students in 1990.

Appendix Figure A4 shows the dynamic effects of university expansion on collaborative patents and industry patents. In both panels, the dashed line presents the estimation results using the number of university teachers in 1990 as the treatment intensity, and the solid line presents the estimation results using the number of university students in 1990 as the treatment intensity. In both panels, there seems to exist an upward trend in outcome variables leading the treatment year of 1999. Starting from 2000, the numbers of both types of patents rose more dramatically, with more pronounced effects in later years.

The deviation in pre-trends across cities with varying treatment intensity raises concerns about potential estimation bias. We investigate and resolve this issue by the following. First, we detect the presence of a trend break by estimating the trend-break model in Equation (1.7), and we report the results in Table 1.4 using the number of university teachers in 1990 as the treatment intensity.[30] Across all columns in the table, we observe statistically significant evidence of trend breaks. The existence of a slope change in the variables of interest suggests the presence of a causal impact of the expansion on university scale and industry patenting activities (Almond et al. 2019).

Next, we strip away city-specific pre-expansion linear time trends following the de-trend approach in Bhuller et al. (2013), Monras (2019), and Garcia-López et al.

[30]The results using the number of university students in 1990 as the treatment intensity is reported in Appendix Table A2.

(2020). We present the corresponding results in Table 1.5. Compared with Table 1.3, the point estimates are very similar for the number of university teachers but reduced a bit for the numbers of different classifications of patents. For example, after adjusting for pre-trends, the estimates imply that cities with 1,000 more university teachers in 1990 experienced an additional increase of 83.89 industry patents on average. Again, the results suggest strong spillovers to local innovation activities from universities. In general, the results show strong robustness.

Furthermore, we re-estimate event study models to verify that the parallel trend assumption is satisfied after we strip away city-specific pre-trends. We report the results in Figures 1.3 and 1.4.[31] We do not observe significant pre-trends in the event study estimates for both figures. This suggests that the extent to which universities expanded at the city level was not predicted by any projected changes in local economic activities in deviation from city-specific time trends. We also find that the impact of university expansion on industry patents is increasing over time, suggesting dynamically increasing spillovers from universities to industry sectors. The increasing effects could be explained by the continually increasing scale of the higher education sector because the university expansion lasted for many years. It is also consistent with the idea that agglomeration spillovers tend to self-amplify once the initial shock takes place.

We conduct two additional sets of robustness checks to further corroborate our findings. First, we follow the parametric event study approach in Dobkin et al. (2018) and estimate the following specification:

$$\begin{aligned} IndustryInno_{c,t} = {} & \mu \times Treatment_c \times \ell + \sum_{\ell=1}^{8} \beta_\ell \times Treatment_c \times \mathbf{1}\{t = 1999 + \ell\} \\ & + \alpha_c + \gamma_t + \varepsilon_{c,t}, \end{aligned} \tag{1.9}$$

where ℓ indicates the year relative to 1999; μ captures the slope of the trend; β_ℓ captures year-specific treatment effect after controlling for city-specific time trend.[32] The rest variables are defined the same as before. We plot the corresponding esti-

[31] The detailed estimation results that are used to draw Figures 3 and 4, and Appendix Figures A3 and A4 are reported in Appendix Tables A3.A and A3.B.

[32] We choose to include linear trends in the model because the non-parametric event study estimates in Figures A3 and A4 display patterns of a linear pre-trend.

mates in Figure 1.5. The dashed lines capture the estimated linear trends. The gap between the crosses and red dashed line represents year-specific treatment effects using the number of university teachers in 1990 as the treatment measure. The gap between the circles and blue dashed line represents year-specific treatment effects using the number of university students in 1990 as the treatment measure. The evidence suggests a large effect of the expansion on industry patents. The patterns are consistent with the lower panel of Figure 1.4. This is not surprising as the parametric event study approach is analogous to the de-trend approach despite small technical variations (Goodman-Bacon 2018, 2021; Rambachan and Roth 2022).

Second, we obtain robust inference using the "honest approach" proposed in Rambachan and Roth (2022) after specifying how different post-treatment violations of parallel trends can be from the pre-treatment differences in trends. Hence, this approach allows us to address potential estimation bias arising from not only the presence of linear trends but also potential deviations from linearity. It also addresses further concerns that pre-trend tests implemented in the event study setup may fail to detect violations of parallel trends due to low statistical power or potential distortions arising from selection (Roth 2022). As in Rambachan and Roth (2022), we assume that the differential trends evolve smoothly over time (smoothness) and that the possible non-linearities in the post-treatment difference in trends are bounded by observed non-linearities in the pre-treatment difference in trends (relative magnitude bounds). To be consistent with Rambachan and Roth (2022), we use δ_t to indicate the difference in trends between the treated and control groups and specify the restriction as follows:

$$\Delta^{SDRM}(\bar{M}) = \left\{ \delta : \forall t \geqslant 0, |(\delta_{t+1} - \delta_t) - (\delta_t - \delta_{t-1})| \leqslant \bar{M} \cdot \max_{s<0} |(\delta_{s+1} - \delta_s) - (\delta_s - \delta_{s-1})| \right\}, \tag{1.10}$$

where Δ is a set of possible differences in trends, and $\bar{M}$ governs the amount by which the slope of δ_t can change after the treatment period. If $\bar{M} = 0$, it requires the trend to be linear, which shares similar ideas as in Equations (1.7) and (1.9). If $\bar{M} > 0$, it means that we further allow a deviation from a linear trend in the post-treatment period, and the maximum deviation is bounded by $\bar{M} \geqslant 0$ times the equivalent maximum in the pre-treatment period.[33] We then construct robust confi-

[33]Applied researchers usually test the null hypothesis $\delta_{pre} = 0$ to assess the existence of the pre-

dence intervals of the treatment effect under the smoothness and relative magnitude bounds assumptions using the R package provided by Rambachan and Roth (2022).

We report the findings in Figure 1.6. The figure presents robust confidence sets for the estimated treatment effect averaged across post-treatment years, under the restrictions of $\Delta^{SDRM}(\bar{M})$. The left panel uses the number of university teachers in 1990 as the treatment intensity, and the right panel uses the number of university students in 1990 as the treatment intensity. The blue confidence intervals are obtained without making adjustments for pre-trends.[34] The red confidence sets depict the set of confidence intervals of the estimated coefficients if we allow for a deviation from a linear trend with the maximum deviation specified in Equation (1.10). We find that, even when we allow $\bar{M} = 0.25$, the estimated causal impact of university expansion on industry patents is still statistically different from zero. The breakdown value for a null effect is around $\bar{M} = 0.5$. Thus, even if we allow further deviations from the pre-expansion linear trend, as long as such deviations are not "too big", we are still able to claim the presence of a causal relationship.

1.5.2 Ring-level Analysis on Knowledge Spillovers

The key focus of this paper is to infer knowledge spillovers from the geographic nature of university spillovers. In this section, we present estimation results from ring-level analyses of the effects of the university expansion on industry patenting activities in close proximity to universities. Specifically, we extend the difference-in-differences framework by further examining whether the impact is larger in areas near universities relative to areas farther away. This within-city variation allows for estimating a triple-differences model, as in Equation (1.2).

Table 1.6 presents the estimation results when we use the number of university teachers in 1990 as the proxy for treatment intensity. The results using the number of university students in 1990 as the treatment proxy are reported in Appendix Table A5. We limit our analysis to areas within 5 km of universities in the baseline regressions. In Columns (1)-(3) of Table 1.6, we report results without removing the pre-expansion linear time trend. In Columns (4)-(6), we report results after re-

treatment non-parallel trends. See Section 2.2 in Rambachan and Roth (2022) for a more detailed discussion.

[34] The point estimate is the average of year-specific estimates of the treatment effect in the post-treatment period.

moving pre-expansion linear time trend.[35] In Columns (1) and (4), we control for city fixed effects, year by ring fixed effects, in addition to treatment by ring dummy interactions. In Columns (2) and (5), we control for year by ring, year by city, and city by ring fixed effects. The latter specification is a standard generalized triple-differences model in which we treat the 4.5-5 km ring as the reference group.

The estimation results suggest that the university expansion significantly increases the number of industry patents in the closest concentric rings and the effects attenuate sharply with geographic distance. While the results are robust and consistent across all specifications, we focus on the results in Column (5), which is our preferred specification. The estimates suggest that cities with 1,000 more university teachers in 1990 experienced an additional increase of 76.5 industry patents in the 0-0.5 km ring relative to the 4.5-5 km ring after the university expansion. This effect reduces to 15.3 in the 0.5-1 km ring, which is smaller by a factor of 5. The effect further reduces as we move to the outer rings and becomes statistically insignificant after the 2 km radius.[36] In Column (6), we divide the coefficient estimates in Column (5) by the average number of patents in the corresponding ring during the pre-expansion period, which provides information on the percentage change of industry patents in each ring because of the university expansion. Again, the attenuation is quite dramatic in percentage terms. In the 0-0.5 km ring, industry patents increased by a factor of 3.12, while in 2-2.5 km ring, industry patents only increased by 43 percent. The attenuation is muted after 2-2.5 km ring.

In Figure 1.7, we plot the dynamic effects of the university expansion on industry patents in different concentric rings after we remove the pre-expansion linear time trend.[37] Panels (a) and (b) use the number of university teachers in 1990 and the number of university students in 1990 as the treatment intensity proxy, respectively. Again, the results show that the impact on the number of patents in the 0-0.5 km ring is the largest, followed by the second ring, third ring, and so on. As will be obvious in this paper, this attenuation pattern is what we consistently find in

[35] The necessity of addressing potential pre-trends is evident in Appendix Figure A5, which shows a small upward trend in the number of industry patents in the nearest ring (ring 1).

[36] To mitigate the concern that many patents are of low quality, we conduct a robustness check in which we restrict our sample to patents with at least one citation. The results are qualitatively similar. We present the results in Appendix Tables A6.A and A6.B.

[37] Appendix Table A7.A and A7.B. present the corresponding regression results.

all specifications. More important, the sharp increasing trend in the 0-0.5 km ring suggests that the long-run benefit of locating near a university could be more amplified than the short-run effects. The figure also shows that the pre-trends are well controlled for, so the parallel trend assumption is not rejected in this case.

Next, we estimate a trend break model to detect the presence of a slope change in the number of industry patents at different distances (rings) as a result of the university expansion, as in Equation (1.8). The results are presented in Table 1.7 when we use the number of university teachers in 1990 as the treatment intensity.[38] Columns (1)-(10) look at each ring separately, and Column (11) pools all rings together and uses 4.5-5 km ring as the reference group. When studying each ring separately, we observe statistically significant evidence of trend breaks in all rings. Column (11) suggests that, comparing with the 4.5-5 km ring, the estimated slope change is statistically significant for the 0-0.5 km, 0.5-1 km, 1-1.5 km, and 1.5-2 km rings. More important, we find that the slope change is much larger in the closer rings. The results suggest that the trajectory of industry patenting activities experienced a trend break because of the expansion and that the causal impact of the expansion follows a dramatic attenuation pattern over space. This pattern supports our previous findings in Table 1.6.

As a robustness check, we present the estimation results from the parametric event study (Dobkin et al. 2018) in Figure 1.8. Panels (a) and (b) present results using the numbers of university teachers and students in 1990 as the proxy for treatment intensity, respectively. To save space, we only present the results for rings 1, 3, 6, and 9. The dashed lines in the figures represent the estimated linear trends in the number of industry patents (μ in Equation (5.1)). The gap between the crosses (circles) and the dashed lines capture the treatment effects of the expansion in deviation from a linear trend. Rings 1 and 6 are represented by crosses; Rings 3 and 9 are represented by circles. The figure shows that there is a significant effect of the university expansion on nearby industry patents in deviation from a linear time trend and that the effect is larger in closer rings to universities, a result we repetitively find in all specifications.

We also obtain robust inference for the ring-level analysis using the "honest ap-

[38] The results using the number of university students in 1990 as the treatment intensity is reported in Appendix Table A8.

proach" proposed in Rambachan and Roth (2022), as the second robustness check. The results are presented in Figure 1.9. The procedure is the same as what we described for the city-level analysis, except now we estimate and present the confidence intervals for each ring separately. The blue confidence intervals are obtained without making adjustments for pre-trends, and the red confidence intervals are obtained when we allow for a deviation from a linear trend with the maximum deviation specified in Equation (1.10). The general patterns are similar to the results in city-level analysis. That is, as long as deviations from a linear trend are not "too big", we continue to find statistically significant evidence of a causal impact of the university expansion on innovation activities. The values contained in the confidence intervals across rings also suggest a sharp attenuation of the impact.

In sum, we document consistent evidence across various specifications suggesting that spillovers from universities are very localized and dissipate sharply with geographic distance. We observe that the largest spillover effects take place within 2 km around the universities, and the impact on the 0-0.5 km ring is more than 30 times larger than that on the 1.5-2 km ring (Column (5) of Table 6). Similar spatial attenuation of agglomeration externalities is also documented in other studies focusing on different settings. For instance, Andersson et al. (2009) shows that between one-third and one-half of the total effect on productivity resulted from a university is within 5 km of the university. Rosenthal and Strange (2008) finds that the effect of urbanization economics on worker productivity is about half as large at distances over 8 km as it is at closer distances. Arzaghi and Henderson (2008) shows that the effect of localization economies on the birth of advertising agencies in Manhattan is mainly within 500 meters. Baum-Snow et al. (2021) finds that revenue and productivity spillovers that operate between firms are within 75 meter to 250 meter radius. As similarly argued in those studies, this important geographic decay of university spillovers suggests that knowledge spillovers play an important role in the effects of universities on local innovation (Arzaghi and Henderson 2008). It would be hard to reconcile such a sharp attenuation pattern with other explanations, such as improved local infrastructure or increased supply of high-skilled labor.[39]

[39] Our results do not exclude the possibility that other mechanisms are present in the neighborhoods of universities. We only claim that, without knowledge spillovers, the impact of university activities should not display a dramatic spatial decay pattern.

Next, we conduct a set of extension and heterogeneity analyses. In the baseline ring-level analysis, we restrict our focus to areas within 5 km of universities. This restriction has two implications when interpreting the estimated coefficients. First, in the triple-differences specification, we use the 4.5-5 km ring as the reference area. Thus, the estimated coefficients capture the impact of the university expansion on the inner rings relative to the impact on the 4.5-5 km ring. Second, the area restriction ignores the possible impact of the university expansion outside the area. In Table 1.8 and Appendix Table A9, we extend the analysis to 10 km around universities, which usually covers a significant share of city areas with innovation activities. The quantitative results on spatial decay patterns are very similar. To better visually reveal the spatial decay pattern of university spillovers, Figures 1.10 and 1.11 present the spatial decay of university spillover benefits using the impact on the 0-0.5 km ring as the reference.[40] We can clearly see the strong spatial decay of the impact, especially within the first 2 km (4 rings) of universities. More important, the spillovers are small and stable beyond this scope, which suggests that we are not missing much by focusing on the 5 km areas around universities.

We further explore how knowledge spillovers from universities interact with other complementary factors, such as industrial and skill composition. In Table 1.9, we explore the heterogeneous effects of the university expansion across different regions.[41] It is well known that the Eastern coastal region of China is the most developed, followed by the Central region, and then the Western region. The industry structure across those regions is quite different. The Eastern region is the most successful in industrial transformation and upgrading, and it comprises high-tech manufacturing concentrations, such as telecommunications and software. The Western region is heavily concentrated with traditional manufacturing industries such as the steel industry. Therefore, spillovers from universities could be different across regions. The estimation results show that the impact of the university expansion is ubiquitous but most pronounced in the Eastern region.

In Table 1.10, we explore the heterogeneous effects of the university expansion

[40]The corresponding regression results are presented in Appendix Tables A10.A and A10.B.

[41]Table 1.9 uses the number of university teachers in 1990 as the proxy for treatment intensity. The results using the number of university students in 1990 as the proxy are reported in Appendix Table A11.

across industries with different human capital intensity.[42] We define the human capital intensity of an industry in the following way. We assign each patent to a two-digit industry based on the reference table of International Patent Classification and National Industries Classification issued by the State Intellectual Property Office of China.[43] We obtain information on the share of workers with a college education and above from the ASIF dataset. Based on this information, we divide industries into high, medium, and low human capital intensity industries, depending on whether the industry-specific college employee ratio belongs to the top, middle, or bottom one-third of the distribution.[44] Finally, we separately count the number of patents linked to the high, medium, and low human capital intensity industries in each concentric ring. The estimation results in Table 1.10 suggest that the spatial attenuation of university spillovers is more pronounced for industries that are more reliant on high-skilled labor.

1.6 Results on Patent Citations and New Products

1.6.1 Patent Citations

We now present direct evidence of knowledge flows from universities to nearby areas by examining the changes in patent citation links because of the university expansion.

In Table 1.11, we examine the effects of university expansion on patent citation links near universities. In Column (1), using the number of university teachers in 1990 as the treatment intensity measure, we examine the impact of university expansion on the number of times when industry patents in different rings cite university patents. We find that university patents are cited by more industry patents near universities after the university expansion. For example, after the university expansion, cities with 1,000 more university teachers in 1990 experienced an additional

[42]Table 1.10 uses the number of university teachers in 1990 as the proxy for treatment intensity. The results using the number of university students in 1990 as the proxy are reported in Appendix Table A12.

[43]The reference table can be found at http://www.sipo.gov.cn/gztz/1132609.htm. It is possible that a patent can be matched with more than one industries, in which case we count this patent in all the industries that it is linked to.

[44]High human capital industries include, for example, the chemical, electrical, and telecommunications industries; medium human capital industries include, for example, the food and beverage industries; and low human capital industries include, for example, the leather and wood industries.

increase of 0.56 times when industry patents in the 0-0.5 km ring cite university patents, relative to that in the 4.5-5 km ring. The effect attenuates fast when we move away from the universities. The corresponding impact is 0.08 in the 0.5-1 km ring relative to the outermost ring, which is smaller than the effect in the 0-0.5 km ring by a factor of 7. The effect decays entirely after the 2 km radius, and the decay speed is as sharp as what we document for the effects on new patents. It is also consistent with the common perception that knowledge spillovers require close-range communications and interactions and, hence, decay fast spatially. In Column (3), we use the number of university students in 1990 as the proxy for treatment intensity, and we find very similar patterns.

A possible argument against the identification of our triple-differences approach is that there may exist unobserved time-varying city- and location (distance)-specific factors that contribute to the increase in location-specific innovation activities after the university expansion. In this case, the increased number of patents that cite university patents in closer locations could result from the scale effect proportional to the increase in the number of total new patents driven by the unobservables. Such a hypothetical scenario is possible but very unlikely given the rare coincidence of multiple co-evolving factors—those factors must have the same timing as the university expansion and systematically impact innovation in a similar spatial pattern. Despite being remotely plausible, we conduct a falsification test to rule out such a possibility.

We examine the impact of university expansion on the spatial nature of the cases where nearby industry patents cite patents far away from universities. If the presence of the unobservables, coupled with the scale effect, forms the underlying mechanism, then we should observe that the impact on the number of cases where nearby patents cite patents far away from universities follows similar spatial decay patterns. Specifically, we examine whether patents outside the 5 km radius of universities are cited more by patents closer to universities after the university expansion. As shown in Columns (2) and (4) of Table 1.11, we do not find a clear spatial decay pattern of the impact. The evidence suggests that our results are not driven by unobserved time-varying city- and ring-specific factors that coincide with the university expansion and that follow a spatial attenuation pattern.

1.6.2 New Products

While patents are informative in measuring innovation, patenting only captures an intermediate step in converting new ideas into economic outputs. In this section, we examine the impact of the university expansion on creation of new products in nearby manufacturing firms by taking advantage of previously under-explored information on firm new products reported in the ASIF. We report the summary statistics of firm characteristics for our final regression sample in Appendix Table A13.

Table 1.12 reports the estimated impact of university expansion on the new product ratio of manufacturing firms in different rings when we use the number of university teachers in 1990 as the proxy for treatment intensity.[45] To capture dynamics, we report the estimated impact separately for the years after 2000, 2002, 2004, and 2006. That is, we report the average impact of university expansion from 2000 to 2007 in Column (1), the average impact of university expansion from 2002 to 2007 in Column (2), and so on. Two patterns emerge. First, for any given column, the impact of university expansion on nearby firms' new product ratio decays as the distance between firms and universities increases. The attenuation pattern is clear, but the speed of attenuation is not as fast as that for patents. Second, the impact gradually increases in later years. This increasing trend is evident when we compare the impact across different columns.

The results in Table 1.12 supplement our analyses on patents by showing that the university expansion also results in an increase in new commercial product sales at nearby firms, which, to some degree, reflects the economic value of innovations. This effect could be explained by a combination of nearby existing firms innovating more and more innovative firms sorting into the neighborhood of universities. Table 1.12 does not intend to distinguish these two channels as they both indicate that there must be some advantages to be in the proximity of universities. The fast decay speed of the impact further suggests that knowledge spillover is a major underlying driving force. We also estimate a specification with firm fixed effects. The results are presented in Appendix Table A15. The coefficients are in general smaller than

[45]The results using the number of university students in 1990 as the treatment intensity are reported in Appendix Table A14.

those in Table 1.12 but the attenuation pattern is still evident. The findings suggest that both the intensive margin and the extensive margin are in effect.

In Table 1.13, we examine the heterogeneous effects of university expansion on new product ratio using the number of university teachers in 1990 as the proxy for treatment intensity.[46] Columns (1)–(3) show the heterogeneous impact across industries with different levels of human capital intensity. The pattern that appears again is the attenuation of the impact over geographic distance. Moreover, we find that potential knowledge spillovers are larger in industries with higher human capital intensity, which is consistent with complementarity between human capital and knowledge spillovers. Columns (4)–(5) explore whether the impact varies for SOEs versus non-SOEs. Evidence suggests that the impact is more pronounced for non-SOE firms, which may be because non-SOEs are in general smaller in size and more productive than SOEs.[47] Thus, they are more active in the market and benefit more from learning and exchanging information.

1.7 Conclusion

Knowledge and innovation play a central role in advancing the technology frontier and promoting economic growth. Yet, despite being the center of knowledge creation and dissemination, the explicit role of universities in contributing to the innovation process is still understudied (Akcigit 2017). This paper exploits a unique quasi-experiment of university expansion in China to study the impact of university activities on local innovation. In particular, we utilize rich geocoded data on patent generations, patent citation links, and new products from firms to examine the geographic nature of the university impact and to identify the role of knowledge spillovers.

We find that the university expansion significantly increases universities' own innovation capacity, which results in a dramatic boom of nearby firms' patenting activities. More important, the impact attenuates sharply with spatial distance. For example, the magnitude of the impact on nearby firm patenting activities reduces

[46]Appendix Table A16 reports the results using the number of university students in 1990 as the treatment intensity.

[47]This result is consistent to Acs et al. (1994). They use new product announcement data from the U.S. Small Business Administration and show that small firms are the recipients of nearby R&D spillovers.

by about 80 percent from 0-0.5 km ring to 0.5-1 km ring around a university. There is another 65 percent decline of the impact when moving from 0.5-1 km ring to 1-1.5 km ring around the university. The result implies significant but very localized knowledge spillovers from universities. Further analysis suggests that the university expansion boosts nearby firms' new products and induces more industry patents to cite university patents. Those effects also follow similar spatial decay patterns. Taken together, these findings unanimously point to the importance of knowledge spillovers in fostering innovation in close proximity to education and research institutions. Thus, the evidence justifies the continually increasing support for research universities as a viable policy instrument for the government to promote long-term economic growth.

While our empirical analysis identifies and highlights the role of knowledge spillovers, future work would benefit from further explorations on the channels through which knowledge spillovers take place in a self-reinforcing way, as suggested by the dynamic evidence that we document in this paper. For instance, to take advantage of increased knowledge spillovers, nearby firms may hire more high-skilled labor and explore its complementarity with knowledge. Increased human capital increases the benefits of knowledge spillovers, which then leads to a self-reinforcing innovation process. Alternatively, increased knowledge spillovers could motivate firms to become more innovative and to enter the proximity of research-oriented universities to better draw on spillover benefits. Their entry and clustering could make it easier to use the knowledge from universities or to generate externalities within the clusters. These channels also reinforce the university spillovers. In a way, spillovers from universities can be viewed as both the "seed" and the "flower" of innovation (Harbison and Myers 1965). Altogether, the specific mechanisms explain the dynamic process through which high-tech clusters form in close proximity to higher education institutions.

References

Acemoglu, D. (1995). Reward Structures and the Allocation of Talent. *European Economic Review*, 39(1):17–33.

Acemoglu, D. (2008). *Introduction to Modern Economic Growth*. Princeton University Press, Princeton, NJ.

Acs, Z. J., Anselin, L., and Varga, A. (2002). Patents and Innovation Counts as Measures of Regional Production of New Knowledge. *Research Policy*, 31(7):1069–1085.

Acs, Z. J. and Audretsch, D. B. (1988). Innovation in Large and Small Firms: An Empirical Analysis. *American Economic Review*, 78(4):678–690.

Acs, Z. J., Audretsch, D. B., and Feldman, M. P. (1994). R&D Spillovers and Recipient Firm Size. *Review of Economics and Statistics*, 76(2):336–340.

Aghion, P., Boustan, L., Hoxby, C., and Vandenbussche, J. (2009). The Causal Impact of Education on Economic Growth: Evidence from U.S. *Brookings Papers on Economic Activity*, pages 1–73.

Aghion, P. and Howitt, P. (1992). A Model of Growth Through Creative Destruction. *Econometrica*, 60(2):323–351.

Akcigit, U. (2017). Economic Growth: the Past, the Present, and the Future. *Journal of Political Economy*, 125(6):1736–1747.

Almond, D., Li, H., and Zhang, S. (2019). Land Reform and Sex Selection in China. *Journal of Political Economy*, 127(2):560–585.

Andersson, R., Quigley, J. M., and Wilhelmsson, M. (2004). University Decentralization as Regional Policy: The Swedish Experiment. *Journal of Economic Geography*, 4(4):371–388.

Andersson, R., Quigley, J. M., and Wilhelmsson, M. (2009). Urbanization, Productivity, and Innovation: Evidence from Investment in Higher Education. *Journal of Urban Economics*, 66(1):2–15.

Andrews, M. (2019). Local Effects of Land Grant Colleges on Agricultural Innovation and Output. *National Bureau of Economic Research*.

Anselin, L., Varga, A., and Acs, Z. (1997). Local Geographic Spillovers between University Research and High Technology Innovations. *Journal of Urban Economics*, 42(3):422–448.

Arzaghi, M. and Henderson, J. V. (2008). Networking off Madison Avenue. *Review of Economic Studies*, 75:1011–1038.

Audretsch, D. B. and Feldman, M. (1996). American Economic Association R&D Spillovers and the Geography of Innovation and Production. *American Economic Review*, 86(3):630–640.

Baum-Snow, N., Gendron-Carrier, N., and Pavan, R. (2021). Local Productivity Spillovers. *Working Paper*.

Bhuller, M., Havnes, T., Leuven, E., and Mogstad, M. (2013). Broadband Internet: An Information Superhighway to Sex Crime? *Review of Economic studies*, 80(4):1237–1266.

Bils, M. and Klenow, P. J. (2000). Does Schooling Cause Growth? *American Economic Review*, 90(5):1160–1183.

Brandt, L., Van Biesebroeck, J., Wang, L., and Zhang, Y. (2017). WTO Accession and Performance of Chinese Manufacturing Firms. *American Economic Review*, 107(9):2784–2820.

Card, D. (1995). Using Geographic Variation in College Proximity to Estimate the Return to Schooling. In *Aspects of Labor Market Behaviour: Essays in Honour of John Vanderkamp*, pages 201–220. University of Toronto Press.

Carlino, G. and Kerr, W. R. (2015). Agglomeration and Innovation. In *Handbook of Regional and Urban Economics*, pages 349–404. Elsevier.

Che, Y. and Zhang, L. (2018). Human Capital, Technology Adoption and Firm Performance: Impacts of China's Higher Education Expansion in the Late 1990s. *Economic Journal*, 128(614):2282–2320.

Combes, P.-P. and Gobillon, L. (2015). The Empirics of Agglomeration Economies. In *Handbook of Regional and Urban Economics*, pages 247–348. Elsevier.

Dobkin, C., Finkelstein, A., Kluender, R., and Notowidigdo, M. J. (2018). The Economic Consequences of Hospital Admissions. *American Economic Review*, 108(2):308–52.

Duflo, E. (2001). Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment. *American economic review*, 91(4):795–813.

Duranton, G. and Puga, D. (2004). Micro-Foundations of Urban Agglomeration Economies. In *Handbook of Regional and Urban Economics*, pages 2063–2117. Elsevier.

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

Feldman, M. P. and Audretsch, D. B. (1999). Innovation in Cities: Science-based Diversity, Specialization, and Localized Competition. *European Economic Review*, 43(2):409–429.

Feldman, M. P. and Kogler, D. F. (2010). Stylized Facts in the Geography of Innovation. In *Handbook of the Economics of Innovation*, pages 381–410. Elsevier.

Figueiredo, O., Guimarães, P., and Woodward, D. (2015). Industry Localization, Distance Decay, and Knowledge Spillovers: Following the Patent Paper Trail. *Journal of Urban Economics*, 89:21–31.

Ganguli, I., Lin, J., and Reynolds, N. (2020). The Paper Trail of Knowledge Spillovers: Evidence from Patent Interferences. *American Economic Journal: Applied Economics*, 12(2):278–302.

Garcia-López, M.-À., Jofre-Monseny, J., Martínez-Mazza, R., and Segú, M. (2020). Do Short-term Rental Platforms Affect Housing Markets? Evidence from Airbnb in Barcelona. *Journal of Urban Economics*, 119:103278.

Glaeser, E. L. and Maré, D. C. (2001). Cities and Skills. *Journal of Labor Economics*, 19(2):316–342.

Goodman-Bacon, A. (2018). Public Insurance and Mortality: Evidence from Medicaid Implementation. *Journal of Political Economy*, 126(1):216–262.

Goodman-Bacon, A. (2021). Difference-in-differences with Variation in Treatment Timing. *Journal of Econometrics*, 225(2):254–277.

Griliches, Z. (1979). Issues in Assessing the Contribution of Research and Development to Productivity Growth. *The Bell Journal of Economics*, pages 92–116.

Grossman, G. and Helpman, E. (1991). *Innovation and Growth in the Global Economy*. MIT Press., Cambridge.

Hall, B. H., Link, A. N., and Scott, J. T. (2003). Universities as Research Partners. *Review of Economics and Statistics*, 85(2):485–491.

Harbison, F. and Myers, C. (1965). *Manpower and Education*. McGrawHil, New York.

Hausman, N. (2022). University Innovation and Local Economic Growth. *The Review of Economics and Statistics*, pages 1–18.

Holmes, T. J. (1999). Localization of Industry and Vertical Disintegration. *Review of Economics and Statistics*, 81(2):314–325.

Jaffe, A. B. (1986). Technological Opportunity and Spillovers of R&D : Evidence from Firms' Patents , Profits, and Market Value. *American Economic Review*, 76(5):984–1001.

Jaffe, A. B. (1989). Real Effects of Academic Research. *American Economic Review*, 79(5):957–970.

Jaffe, A. B., Trajtenberg, M., and Henderson, R. (1993). Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations. *Quarterly Journal of Economics*, 108(3):577–598.

Jefferson, G., Hu, A. G., Guan, X., and Yu, X. (2003). Ownership, Performance, and Innovation in China's Large and Medium-Size Industrial Enterprise Sector. *China Economic Review*, 14(1):89–113.

Kantor, S. and Whalley, A. (2014). Knowledge Spillovers from Research Universities: Evidence from Endowment Value Shocks. *Review of Economics and Statistics*, 96(1):171–188.

Kantor, S. and Whalley, A. (2019). Research Proximity and Productivity: Long-term Evidence from Agriculture. *Journal of Political Economy*, 127(2):819–854.

Li, H., Ma, Y., Meng, L., Qiao, X., and Shi, X. (2017). Skill Complementarities and Returns to Higher Education: Evidence from College Enrollment Expansion in China. *China Economic Review*, 46:10–26.

Li, J., Li, L., and Liu, S. (2022). Attenuation of Agglomeration Economies : Evidence from the Universe of Chinese Manufacturing Firms. *Journal of Urban Economics*, page 103458.

Liu, S. (2015). Spillovers from Universities: Evidence from the Land-grant Program. *Journal of Urban Economics*, 87:25–41.

Lu, J. and Tao, Z. (2009). Trends and Determinants of China's Industrial Agglomeration. *Journal of Urban Economics*, 65(2):167–180.

Lucas, R. E. (1988). On the Mechanics of Economic Development. *Journal of Monetary Economics*, 22:3–42.

Marshall, A. (1890). *Principles of Economics*. Mcmillan, London.

Monras, J. (2019). Minimum Wages and Spatial Equilibrium: Theory and Evidence. *Journal of Labor Economics*, 37(3):853–904.

Moretti, E. (2004). Workers' Education, Spillovers, and Productivity: Evidence from Plant-level Production Functions. *American Economic Review*, 94(3):656–690.

Moretti, E. (2021). The Effect of High-Tech Clusters on the Productivity of Top Inventors. *American Economic Review*, 111(10):3328–75.

Murata, Y., Nakajima, R., Okamoto, R., and Tamura, R. (2014). Localized Knowledge Spillovers and Patent Citations: A Distance-Based Approach. *Review of Economics and Statistics*, 96(4):710–728.

Pakes, A. and Griliches, Z. (1980). Patents and R&D at the Firm Level: A First Report. *Economics letters*, 5(4):377–381.

Rambachan, A. and Roth, J. (2022). A More Credible Approach to Parallel Trends. *Working Paper*.

Redding, S. (1996). The Low-Skill, Low-Quality Trap: Strategic Complementarities Between Human Capital and R&D. *Economic Journal*, 106:458–470.

Romer, P. M. (1986). Increasing Returns and Long-Run Growth. *Journal of Political Economy*, 94(5):1002–1037.

Romer, P. M. (1990). Endogenous Technological Change. *Journal of Political Economy*, 98(5):S71–S102.

Rong, Z. and Wu, B. (2020). Scientific Personnel Reallocation and Firm Innovation: Evidence from China's College Expansion. *Journal of Comparative Economics*, pages 1–20.

Rosenthal, S. S. and Strange, W. C. (2001). The Determinants of Agglomeration. *Journal of Urban Economics*, 50(2):191–229.

Rosenthal, S. S. and Strange, W. C. (2003). Geography, Industrial Organization, and Agglomeration. *Review of Economics and Statistics*, 85(May):377–393.

Rosenthal, S. S. and Strange, W. C. (2004). Evidence on the Nature and Sources of Agglomeration Economies. In *Handbook of Regional and Urban Economics*, pages 2119–2177. Elsevier.

Rosenthal, S. S. and Strange, W. C. (2005). The Geography of Entrepreneurship in the New York Metropolitan Area. *Federal Reserve Bank of New York Economic Policy Review*, 11:29–53.

Rosenthal, S. S. and Strange, W. C. (2008). The Attenuation of Human Capital Spillovers. *Journal of Urban Economics*, 64(2):373–389.

Rosenthal, S. S. and Strange, W. C. (2020). How Close is Close? The Spatial Reach of Agglomeration Economies. *Journal of Economic Perspectives*, 34(3):27–49.

Roth, J. (2022). Pre-test with Caution: Event-Study Estimates after Testing for Parallel Trends. *American Economic Review*.

Schofer, E. and Meyer, J. W. (2005). The Worldwide Expansion of Higher Education in the Twentieth Century. *American Sociological Review*, 70(6):898–920.

Thompson, P. (2006). Patent Citations and the Geography of Knowledge Spillovers : Evidence from Inventor- and Examiner-Added Citations. *Review of Economics and Statistics*, 88(2):383–388.

Thompson, P. and Fox-Kean, M. (2005). Patent Citations and the Geography of Knowledge Spillovers : A Reassessment. *American Economic Review*, 95(1):450–460.

Tombe, T. and Zhu, X. (2019). Trade, Migration, and Productivity: A Quantitative Analysis of China. *American Economic Review*, 109(5):1843–1872.

Valero, A. and Van Reenen, J. (2019). The Economic Impact of Universities: Evidence from Across the Globe. *Economics of Education Review*, 68:53–67.

Zhao, S. (2018, November 28). Chinese Silicon Valley – Story of Zhongguancun. *Economic Daily (China)*, page 13.

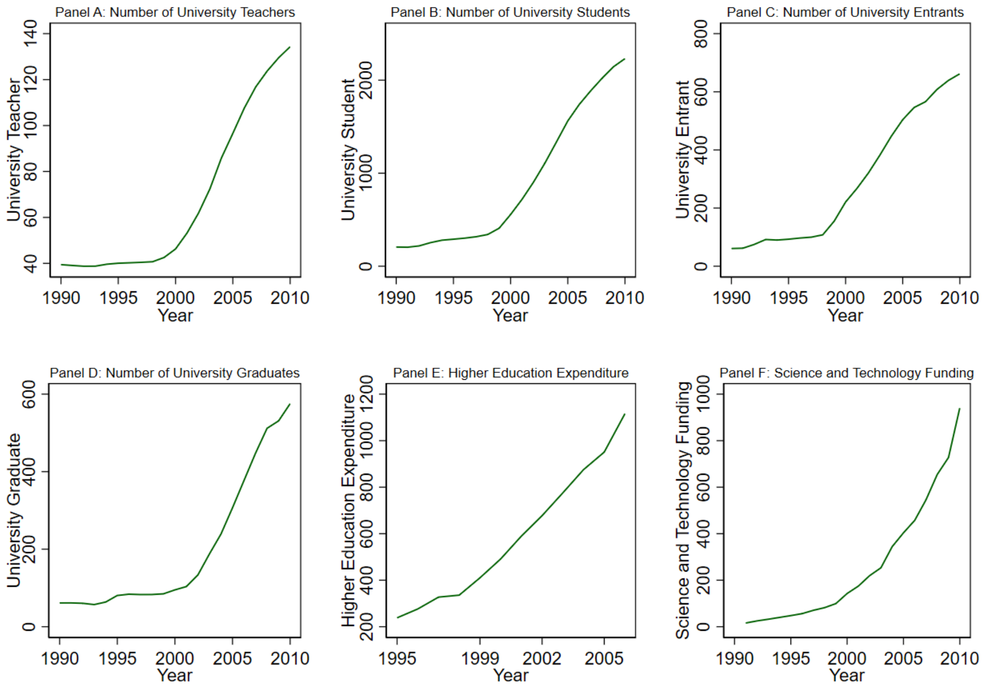


Figure 1.1: Various Aspects of the University Expansion

Notes: Numbers are counted in 10,000 from Panels A to D, and in 100,000,000 yuan in Panels E and F. Data for the numbers of university teachers, university students, university entrants, and university graduates are obtained from the Educational Statistics Yearbook of China. Data for higher education expenditure are from the Educational Finance Statistical Yearbook of China. Data for science and technology funding in the higher education sector are from the Compilation of Statistical Data on University Science and Technology Resource.

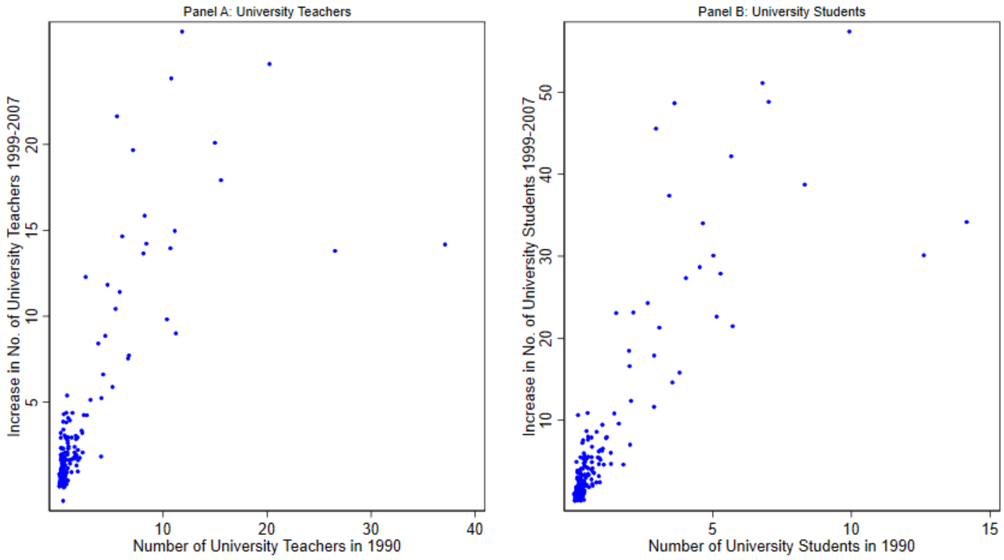
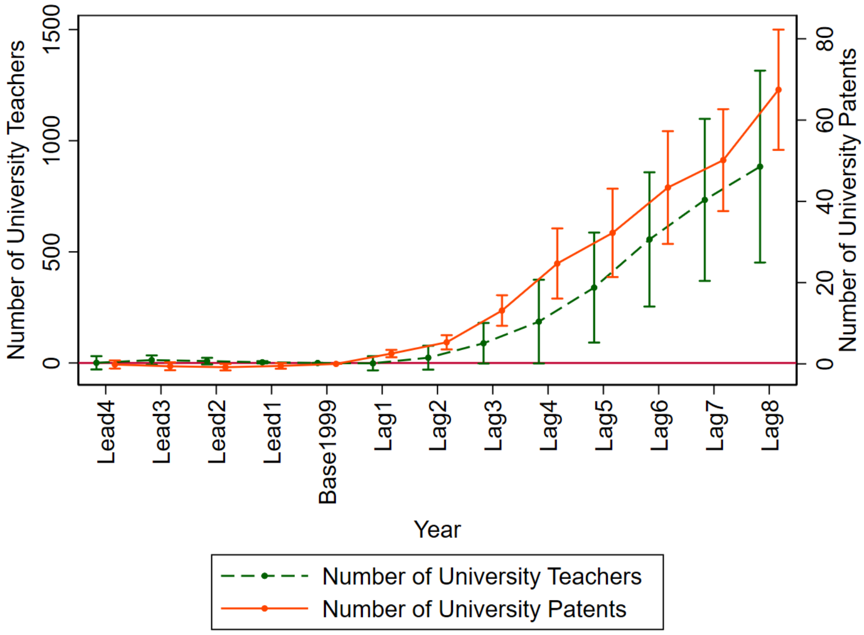
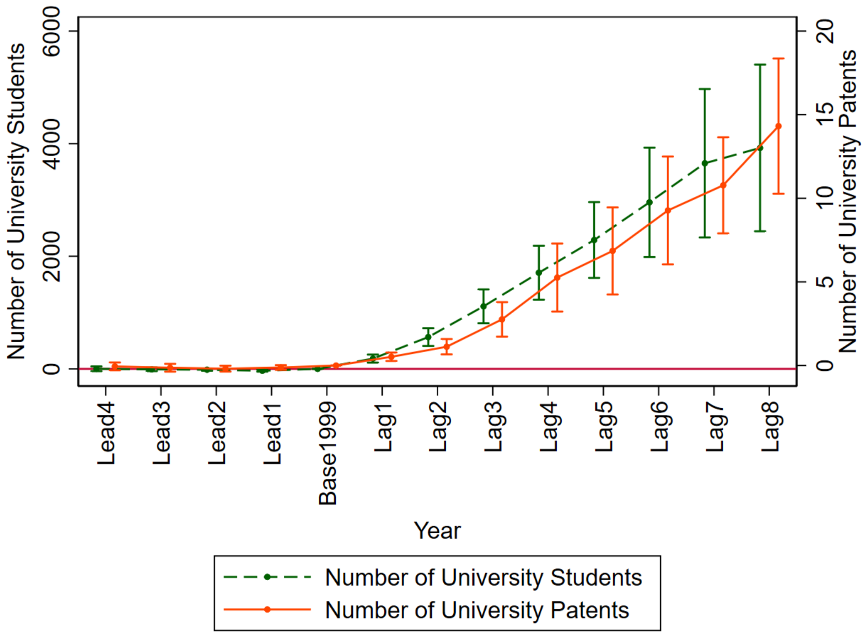


Figure 1.2: Growth in the University Scale between 1999 and 2007 in Relation to the University Scale in 1990

Notes: The number of university teachers is counted in 1,000. The number of university students is counted in 10,000. Data are obtained from the China City Statistical Yearbook.

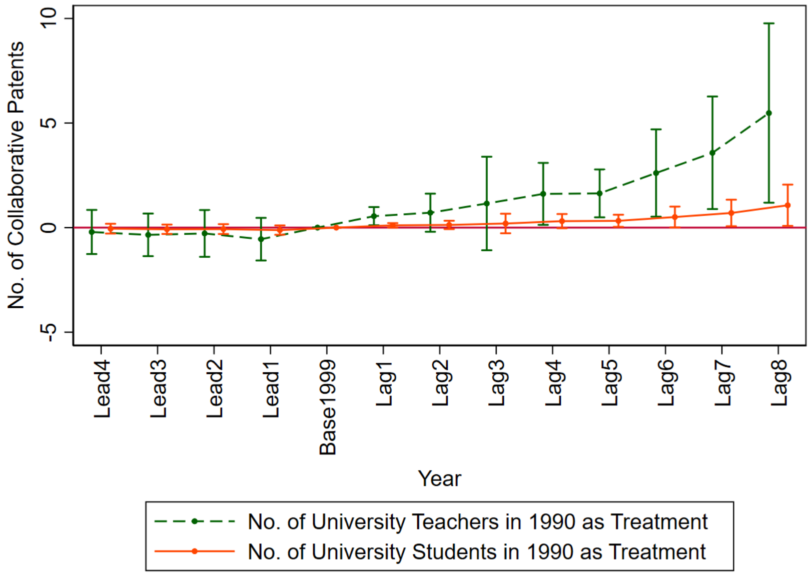


(a) No. of University Teachers in 1990 as Treatment

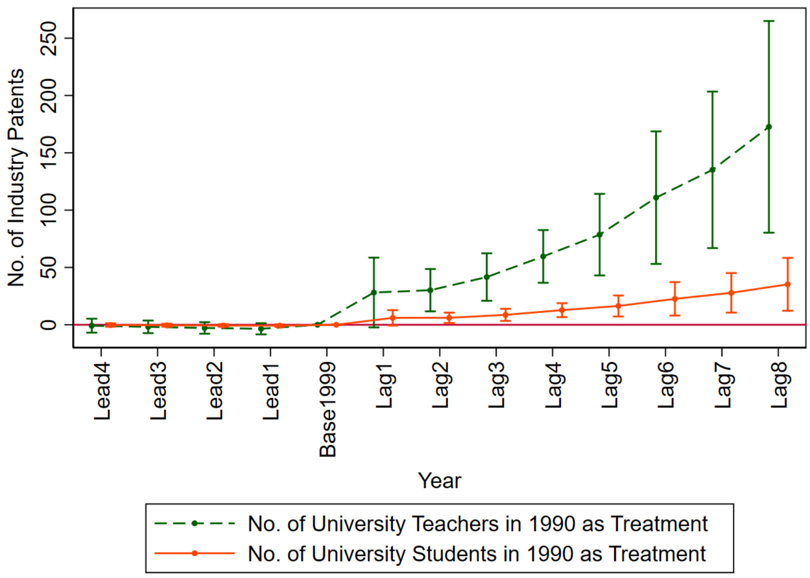


(b) No. of University Students in 1990 as Treatment

Figure 1.3: The Dynamic Effects of University Expansion on the Numbers of University Teachers, Students, and University Patents



(a) Collaborative Patents



(b) Industry Patents

Figure 1.4: The Dynamic Effects of University Expansion on the Numbers of Collaborative Patents and Industry Patents

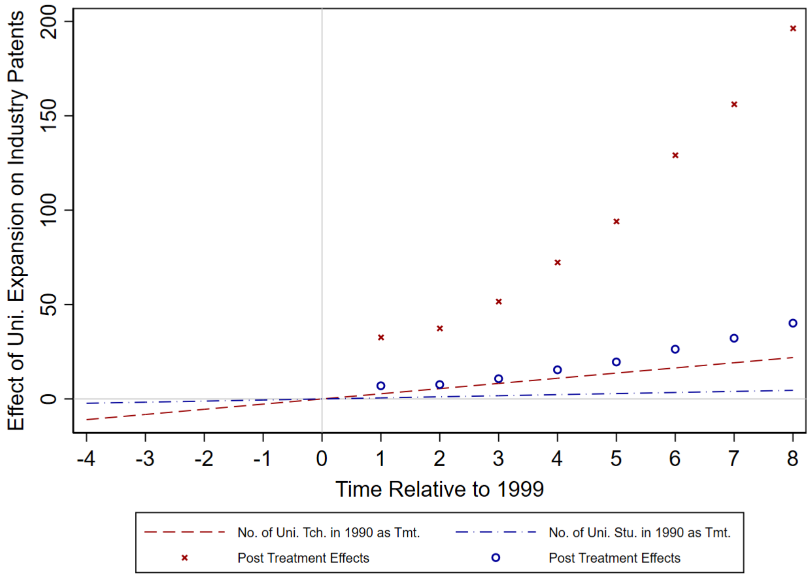


Figure 1.5: The Dynamic Effects of University Expansion on Industry Patents — Parametric Event Study Approach in Dobkin et al. (2018)

Notes: This figure reports the results estimating Equation (1.9), which is a parametric event study approach introduced in Dobkin et al. (2018). The dashed line in the figure represents the estimated linear trend (the corresponding slope is μ in Equation (1.9)). The gap between the crosses (circles) and the dashed lines capture the estimated effects of the expansion.

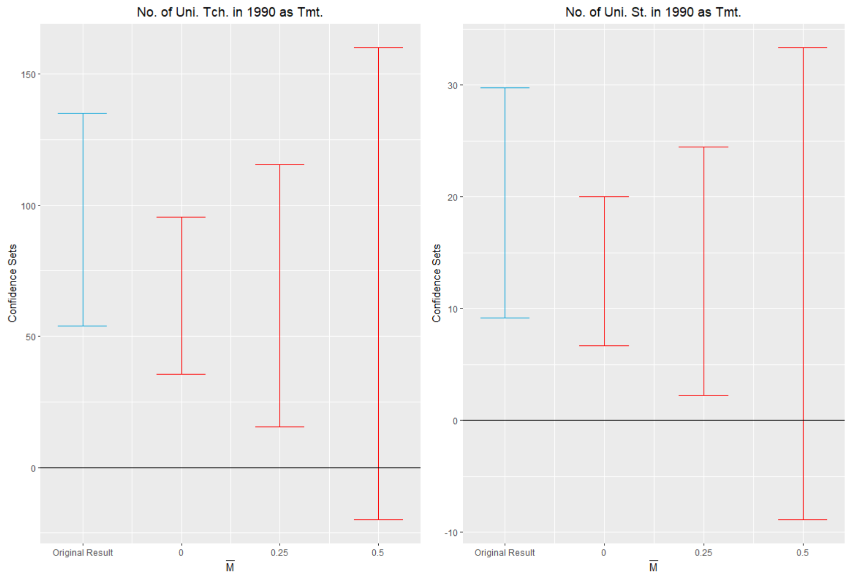
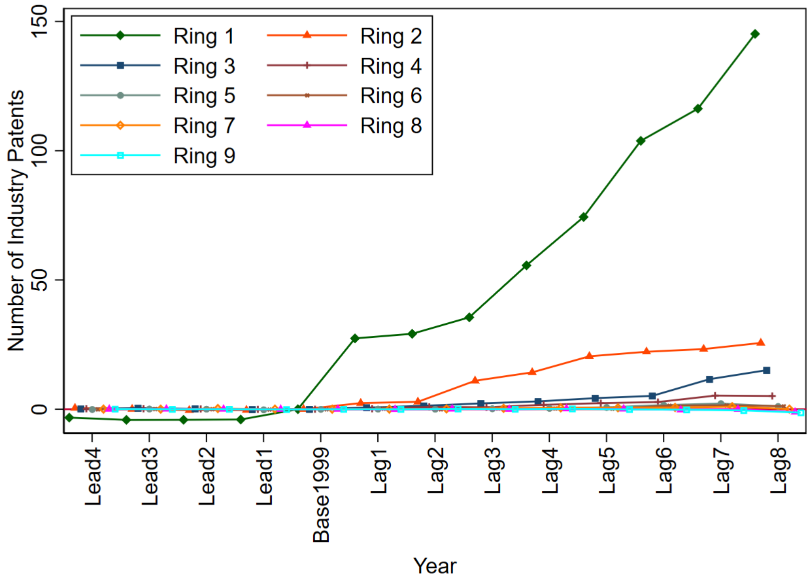
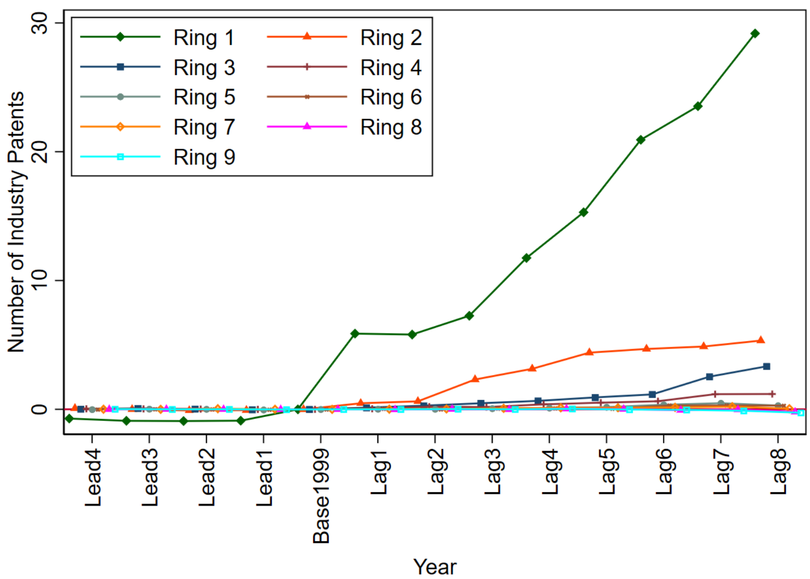


Figure 1.6: “Honest" Approach — Confidence Sets for the Effects of University Expansion on Industry Patents

Notes: This figure reports robust confidence sets for the average treatment effect across all post-treatment periods. It is produced using the R package provided by Rambachan and Roth (2022). The number of university teachers in 1990 is used as the treatment intensity in the left panel, and the number of university students in 1990 is used as the treatment intensity in the right panel. The blue confidence intervals are obtained without making adjustments for pre-trends. The red confidence sets depict the set of confidence intervals of the estimated coefficients if we allow for a deviation from a linear trend as specified in Equation (1.10).

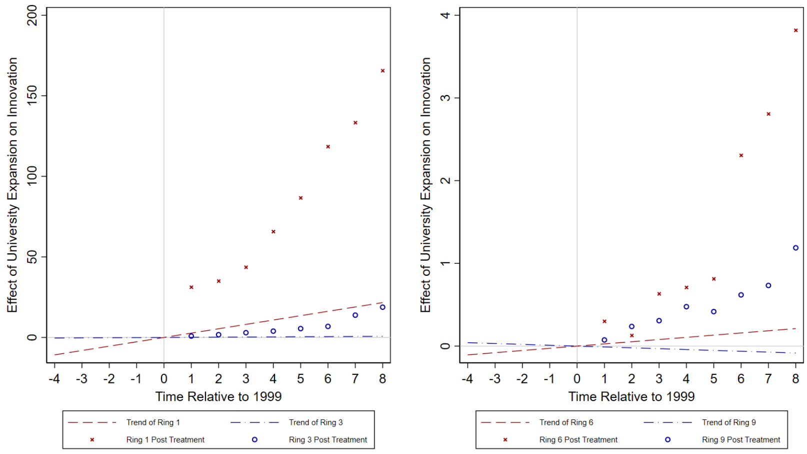


(a) No. of University Teachers in 1990 as Treatment

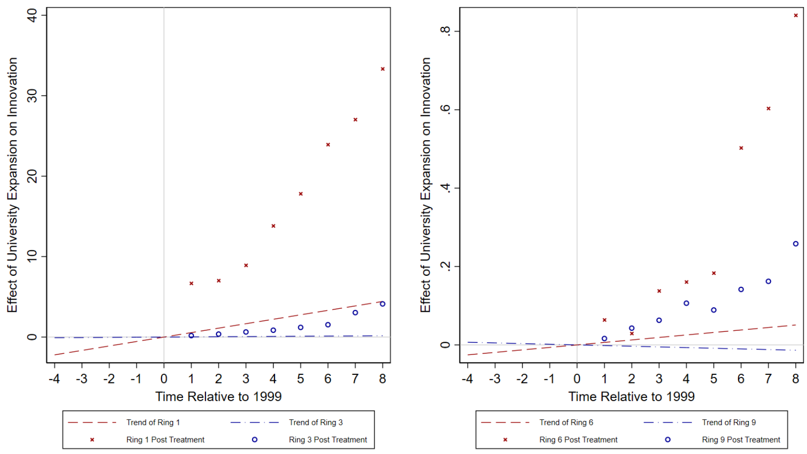


(b) No. of University Students in 1990 as Treatment

Figure 1.7: The Dynamic Effects of University Expansion on the Number of Industry Patents at the Ring Level — Pre-expansion Time Trend Removed



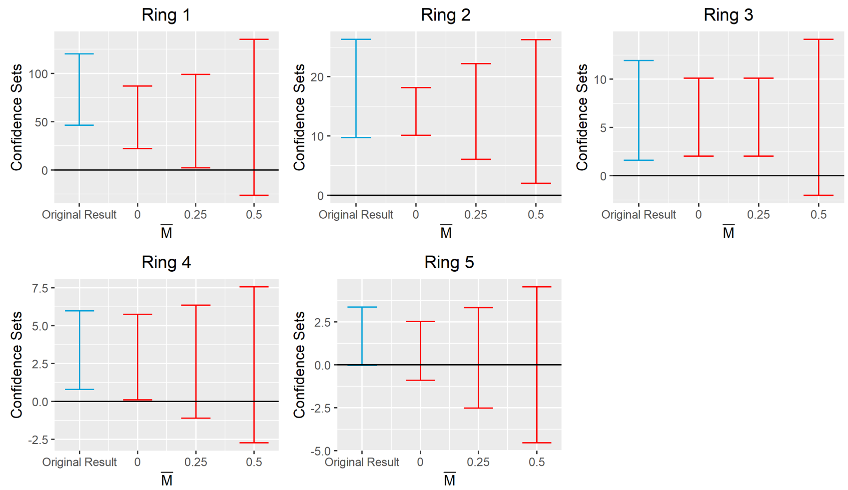
(a) No. of University Teachers in 1990 as Treatment



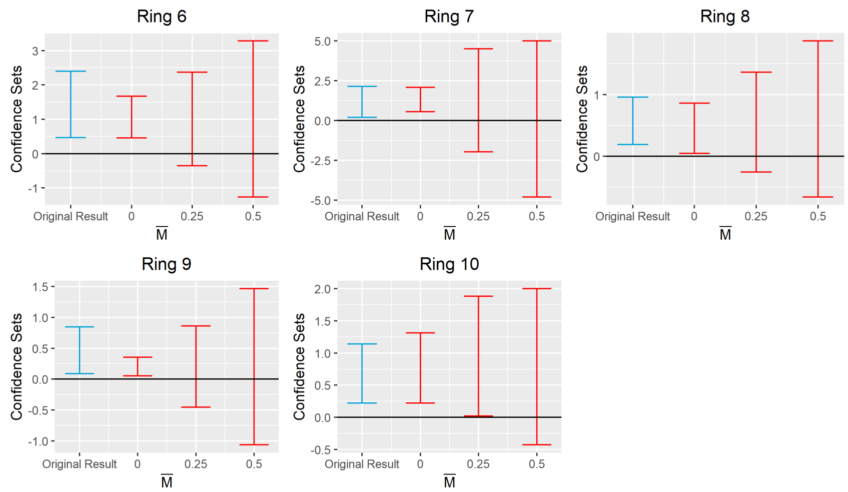
(b) No. of University Students in 1990 as Treatment

Figure 1.8: The Dynamic Effects of University Expansion on Industry Patents at the Ring Level
— Parametric Event Study Approach in Dobkin et al. (2018)

Notes: The specification used for each ring is specified in Equation (5.1), where ℓ is the year relative to 1999. We only present the results for rings 1, 3, 6, and 9 to save space. The dashed lines in the figures represent the estimated linear trends (the corresponding slope is μ in Equation (5.1)). The gap between the crosses (circles) and the dashed lines capture the estimated effects of the expansion. Rings 1 and 6 are represented by crosses; Rings 3 and 9 are represented by circles.



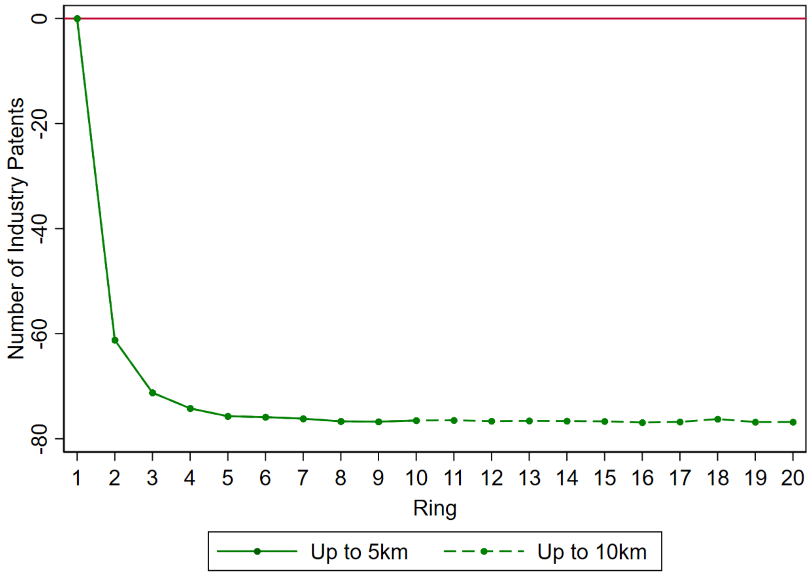
(a) Ring 1 - Ring 5



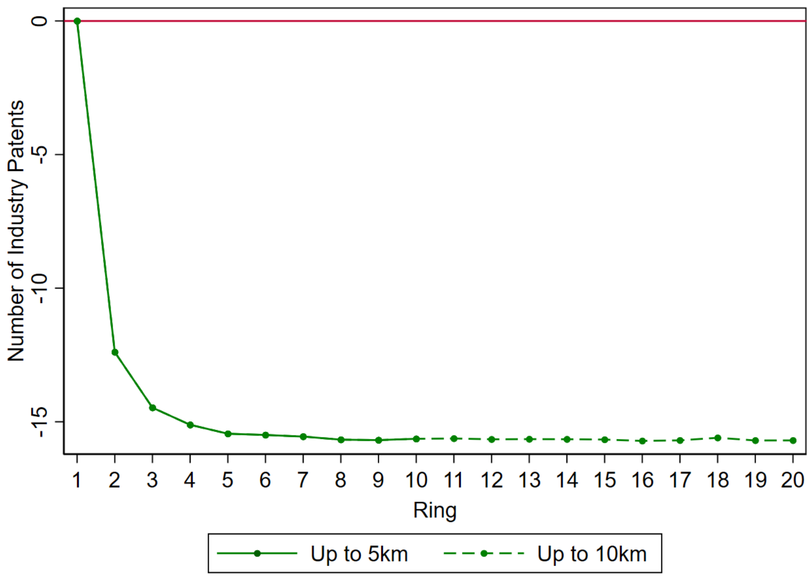
(b) Ring 6 - Ring 10

Figure 1.9: “Honest" Approach — Confidence Sets for the Effect of University Expansion on Patents in Ring 1-10

Notes: This figure reports robust confidence sets for the average treatment effect across all post-treatment periods for each ring. It is produced using the R package provided by Rambachan and Roth (2022). The number of university teachers in 1990 is counted in 1,000, and it is used as the measure of treatment intensity. The blue confidence intervals are obtained without making adjustments for pre-trends. The red confidence sets depict the set of confidence intervals of the estimated coefficients if we allow for a deviation from a linear trend as specified in Equation (1.10).



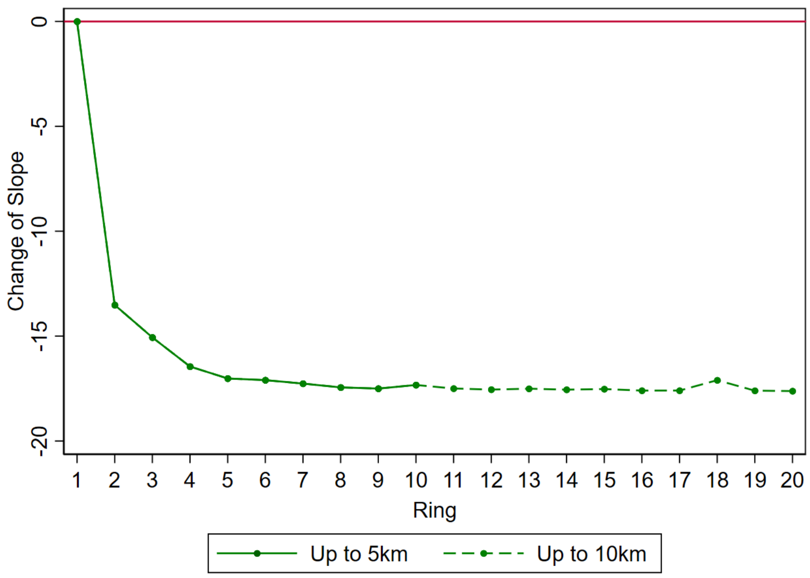
(a) No. of University Teachers in 1990 as Treatment



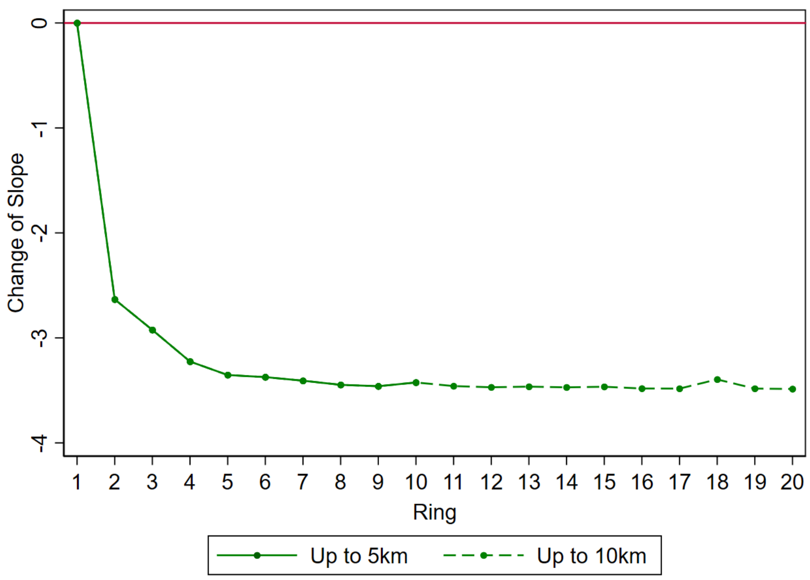
(b) No. of University Students in 1990 as Treatment

Figure 1.10: Spatial Decay of University Spillovers — Relative to the Effect on Ring 1

Notes: This figure depicts the effect of the university expansion on different rings using the de-trend method, relative to the effect on ring 1 (0-0.5 km ring). The number of university teachers (students) in 1990 is used as the measure of treatment intensity in the top (bottom) panel. Both variables are counted in 1,000.



(a) No. of University Teachers in 1990 as Treatment



(b) No. of University Students in 1990 as Treatment

Figure 1.11: Spatial Decay of University Spillovers — Relative to the Effect on Ring 1 (Trend Break Model)

Notes: This figure depicts the effect of the university expansion on different rings using the trend break model, relative to the effect on ring 1 (0-0.5 km ring). The number of university teachers (students) in 1990 is used as the measure of treatment intensity in the top (bottom) panel. Both variables are counted in 1,000.

Table 1.1: City-Level Summary Statistics

Year	(1) Cities	(2) University Teachers	(3) University Students	(4) Total Patents	(5) University Patents	(6) Collaborative Patents	(7) Industry Patents
1995	184	2066.14	14544.88	44.11	2.56	0.30	41.26
1996	184	2092.56	15274.52	49.84	2.86	0.40	46.58
1997	184	2106.44	15928.52	53.55	2.80	0.54	50.21
1998	184	2092.70	16781.35	59.97	3.80	0.56	55.60
1999	184	2111.27	18066.50	71.42	4.97	1.13	65.32
2000	184	2199.40	21705.90	114.96	8.28	1.74	104.93
2001	184	2376.25	28130.00	139.22	12.48	2.23	124.51
2002	184	2648.45	37005.33	197.61	22.43	2.93	172.24
2003	184	3084.78	46802.86	274.86	38.29	4.01	232.56
2004	184	3653.77	58002.03	308.14	49.88	4.65	253.61
2005	184	4310.27	69309.45	404.68	68.80	6.32	329.57
2006	184	4896.28	81839.51	528.77	85.07	8.46	435.23
2007	184	5403.83	88653.08	644.54	105.72	10.63	528.19

Notes: Column (1) reports the number of cities in each year. Columns (2)–(7) report the mean of the respective city-level variable. University patents are the patents filed solely by inventors affiliated with higher-education institutions. Collaborative patents are the patents jointly filed by inventors affiliated with universities and inventors from the private sector. Industry patents are the patents filed solely by inventors from the private sector.

Table 1.2: The Average Number of Patents at the Ring Level

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Year	Ring 1	Ring 2	Ring 3	Ring 4	Ring 5	Ring 6	Ring 7	Ring 8	Ring 9	Ring 10
1995	18.20	6.33	3.40	2.32	1.35	0.80	0.62	0.58	0.46	0.33
1996	20.77	6.64	3.66	2.34	1.64	1.24	0.70	0.66	0.39	0.40
1997	23.45	6.61	3.93	2.23	1.80	1.22	0.85	0.49	0.48	0.46
1998	26.47	7.81	4.25	2.57	2.10	1.44	0.97	0.70	0.43	0.48
1999	33.86	8.82	4.62	2.90	2.42	1.52	1.05	0.77	0.58	0.52
2000	65.55	12.48	6.60	3.89	2.73	1.92	1.37	0.82	0.83	0.77
2001	79.36	14.72	7.93	5.11	3.02	1.97	1.78	1.36	0.88	0.88
2002	107.49	28.36	10.57	6.20	4.92	3.30	2.64	1.71	1.30	1.24
2003	150.23	39.66	13.91	8.51	6.22	4.23	3.51	2.57	2.21	1.61
2004	167.61	44.65	16.84	10.53	6.42	4.56	3.99	3.30	2.39	2.27
2005	211.91	48.41	22.48	13.95	10.19	7.98	6.15	4.27	3.01	3.91
2006	267.29	53.45	32.04	21.36	14.08	11.53	8.41	6.78	4.24	6.33
2007	316.14	65.49	39.30	25.49	17.35	15.03	11.33	8.41	6.40	9.64

Notes: This table reports the average numbers of patents in different concentric rings in each year across cities. Ring i refers to the concentric ring area between the buffer zones $(i-1)$ and i, and the boundaries of consecutive buffer zones are 500 meters apart.

Table 1.3: Impact of University Expansion on University Scale and Innovation — City-level Regression

Dependent Variable: No. of	(1) University Teachers	(2) Total Patents	(3) University Patents	(4) Collabo-rative Patents	(5) Industry Patents
Treatment × After	341.02***	138.74***	33.32***	3.67***	101.75***
	(3.23)	(5.20)	(7.57)	(3.95)	(4.52)
Observations	2384	2392	2392	2392	2392
Year FE	Yes	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes	Yes
Dependent Variable Mean	3006.14	222.44	31.38	3.38	187.68
Adj. R^2	0.913	0.585	0.647	0.676	0.539

Notes: This table reports the estimates of the effects of university expansion on the numbers of university teachers and different classifications of patents. The number of university teachers in 1990 is counted in 1,000, and it is used as the measure of treatment intensity. The number of university teachers is considered as a proxy for university scale. t statistics based on clustered standard errors at the city level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.4: Impact of University Expansion on University Scale and Innovation — City-level Analysis of Trend Break Model

	(1)	(2)	(3)	(4)	(5)
Dependent Variable: No. of	University Teachers	Total Patents	University Patents	Collabo-rative Patents	Industry Patents
Treatment $\times$ Trend	155.40***	32.55***	10.09***	0.647**	21.81***
$\times$After 2000	(4.64)	(5.59)	(9.15)	(2.47)	(4.23)
Observations	2384	2392	2392	2392	2392
Year FE	Yes	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes	Yes
Dependent Variable Mean	3006.14	222.44	31.38	3.38	187.68
Adj. R^2	0.949	0.679	0.873	0.783	0.602

Notes: This table reports the estimates of the slope change in the numbers of university teachers and different classifications patents as a result of the university expansion, using the specification in Equation (1.7). The number of university teachers in 1990 is counted in 1,000, and it is used as the measure of treatment intensity. t statistics based on clustered standard errors at the city level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.5: Impact of University Expansion on University Scale and Innovation — City-level Regression with Pre-expansion Linear Trend Removed

Dependent Variable: No. of	(1) University Teachers	(2) Total Patents	(3) University Patents	(4) Collabo-rative Patents	(5) Industry Patents
Treatment × After	346.06***	116.61***	30.27***	2.44***	83.89***
	(3.28)	(4.37)	(6.88)	(2.63)	(3.73)
Observations	2384	2392	2392	2392	2392
Year FE	Yes	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes	Yes
Dependent Variable Mean	831.91	138.63	24.12	1.24	113.27
Adj. R^2	0.710	0.475	0.589	0.492	0.426

Notes: This table reports the estimates of the effects of university expansion on the numbers of university teachers and different classifications of patents. The city-specific pre-expansion linear trend is removed for the dependent variable in the specifications. The number of university teachers in 1990 is counted in 1,000, and it is used as the measure of treatment intensity. The number of university teachers is considered as a proxy for university scale. t statistics based on clustered standard errors at the city level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.6: Impact of University Expansion on Industry Innovation — Ring-level Regressions

	Pre-expansion Time Trend Not Removed			Pre-expansion Time Trend Removed		
Dependent Variable	Number of Patents					
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment ×	90.41***	89.70***	3.65	76.93***	76.52***	3.12
After × 0.5km	(4.46)	(4.44)		(3.80)	(3.79)	
Treatment ×	19.03***	18.33***	2.53	15.72***	15.31***	2.11
After × 1km	(4.28)	(4.27)		(3.54)	(3.57)	
Treatment ×	6.98**	6.28**	1.58	5.70**	5.29**	1.33
After × 1.5km	(2.55)	(2.42)		(2.08)	(2.04)	
Treatment ×	3.48***	2.78**	1.13	2.71**	2.30*	0.93
After × 2km	(2.60)	(2.33)		(2.03)	(1.93)	
Treatment ×	1.76**	1.06	0.57	1.21	0.81	0.43
After × 2.5km	(1.98)	(1.40)		(1.36)	(1.07)	
Treatment ×	1.49***	0.79*	0.64	1.05*	0.65	0.52
After × 3km	(2.71)	(1.90)		(1.91)	(1.55)	
Treatment ×	1.15**	0.44	0.53	0.75	0.34	0.41
After × 3.5km	(2.13)	(1.04)		(1.39)	(0.80)	
Treatment ×	0.60***	-0.11	-0.17	0.24	-0.16	-0.26
After × 4km	(2.75)	(-0.65)		(1.12)	(-1.01)	
Treatment ×	0.49**	-0.22	-0.46	0.18	-0.23	-0.48
After × 4.5km	(2.14)	(-1.31)		(0.79)	(-1.38)	
Treatment ×	0.70***	-	-	0.41*	-	-
After × 5km	(3.12)	-		(1.81)	-	
Observations	23920	23920	-	23920	23920	-
Treatment × Ring Dummies	Yes	No	-	Yes	No	-
City FE	Yes	No	-	Yes	No	-
Year × Ring FE	Yes	Yes	-	Yes	Yes	-
Year × City FE	No	Yes	-	No	Yes	-
City × Ring FE	No	Yes	-	No	Yes	-
Dependent Variable Mean	18.21	18.21	-	10.91	10.91	-
Adj. R^2	0.387	0.570	-	0.255	0.482	-

Notes: This table reports the estimated effects of university expansion on industry patents at different distances (rings). The city-ring-specific pre-expansion time trend is removed for the dependent variable in Columns (4)-(5). Columns (3) and (6) are obtained by dividing the coefficients in Columns (2) and (5) by the average number of patents in the corresponding ring during the pre-expansion periods, respectively. The number of university teachers in 1990 is counted in 1,000, and it is used as the measure of treatment intensity. t statistics based on clustered standard errors at the city level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.7: Impact of University Expansion on Industry Innovation — Ring-level Regressions of Trend Break Model

	(1) Ring 1	(2) Ring 2	(3) Ring 3	(4) Ring 4	(5) Ring 5	(6) Ring 6	(7) Ring 7	(8) Ring 8	(9) Ring 9	(10) Ring 10	(11) Ring 1 - 10
Treatment $\times$ Trend	17.65^{***}										17.33^{***}
$\times$After$\times$ 0.5km	(3.54)										(3.47)
Treatment $\times$ Trend		4.137^{***}									3.815^{***}
$\times$After$\times$ 1km		(4.66)									(4.56)
Treatment $\times$ Trend			2.588^{**}								2.267^{*}
$\times$After$\times$ 1.5km			(2.02)								(1.85)
Treatment $\times$ Trend				1.202^{***}							0.881^{**}
$\times$After$\times$ 2km				(2.74)							(2.29)
Treatment $\times$ Trend					0.633^{*}						0.311
$\times$After$\times$ 2.5km					(1.90)						(1.10)
Treatment $\times$ Trend						0.554^{**}					0.233
$\times$After$\times$ 3km						(2.60)					(1.41)
Treatment $\times$ Trend							0.388^{**}				0.0670
$\times$After$\times$ 3.5km							(2.23)				(0.50)
Treatment $\times$ Trend								0.208^{**}			-0.114
$\times$After$\times$ 4km								(2.49)			(-1.41)
Treatment $\times$ Trend									0.150^{*}		-0.171^{*}
$\times$After$\times$ 4.5km									(1.80)		(-1.96)
Treatment $\times$ Trend										0.321^{***}	
$\times$After$\times$ 5km										(3.07)	
Observations	2392	2392	2392	2392	2392	2392	2392	2392	2392	2392	23920
City FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Year $\times$ Ring FE	No	No	No	No	No	No	No	No	No	No	Yes
Year $\times$ City FE	No	No	No	No	No	No	No	No	No	No	Yes
City $\times$ Ring FE	No	No	No	No	No	No	No	No	No	No	Yes
Dependent Variable Mean	114.50	26.42	13.04	8.26	5.71	4.37	3.34	2.49	1.82	2.22	18.21
Adj. R^2	0.606	0.672	0.561	0.621	0.450	0.455	0.450	0.379	0.430	0.213	0.627

Notes: This table reports the estimates of the slope change in the number of industry patents at different distances (rings) as a result of the university expansion, using the specification in Equation (3.8). The number of university teachers in 1990 is counted in 1,000, and it is used as the measure of treatment intensity. The trend-break model is used in all specifications. t statistics based on clustered standard errors at the city level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.8: Robustness Check — Ring-level Regressions up to 10 km

	Pre-expansion Time Trend Not Removed		Pre-expansion Time Trend Removed	
Dependent Variable	Number of Patents			
	(1)	(2)	(3)	(4)
Treatment ×	90.41***	90.29***	77.06***	77.07***
After × 0.5km	(4.46)	(4.46)	(3.80)	(3.81)
Treatment ×	19.03***	18.92***	15.85***	15.86***
After × 1km	(4.28)	(4.30)	(3.57)	(3.60)
Treatment ×	6.98**	6.87**	5.83**	5.84**
After × 1.5km	(2.55)	(2.55)	(2.13)	(2.17)
Treatment ×	3.48***	3.37***	2.84**	2.85**
After × 2km	(2.60)	(2.61)	(2.12)	(2.21)
Treatment ×	1.76**	1.65*	1.35	1.36
After × 2.5km	(1.98)	(1.96)	(1.51)	(1.61)
Treatment ×	1.49***	1.38***	1.18**	1.20**
After × 3km	(2.71)	(2.74)	(2.15)	(2.37)
Treatment ×	1.15**	1.03**	0.88	0.89*
After × 3.5km	(2.13)	(2.07)	(1.63)	(1.78)
Treatment ×	0.60***	0.49***	0.37*	0.39**
After × 4km	(2.75)	(2.87)	(1.73)	(2.29)
Treatment ×	0.49**	0.38**	0.31	0.32*
After × 4.5km	(2.14)	(2.14)	(1.37)	(1.84)
Treatment ×	0.70***	0.59***	0.54**	0.55***
After × 5km	(3.12)	(3.07)	(2.39)	(2.86)
Treatment ×	0.44***	0.33***	0.28**	0.30***
After × 5.5km	(3.25)	(3.40)	(2.09)	(3.05)
Treatment ×	0.29**	0.18*	0.14	0.15
After × 6km	(2.06)	(1.75)	(0.97)	(1.45)
Treatment ×	0.34***	0.23***	0.20**	0.21***
After × 6.5km	(3.46)	(3.94)	(2.04)	(3.63)
Treatment ×	0.31**	0.20	0.17	0.18
After × 7km	(2.32)	(1.59)	(1.27)	(1.45)
Treatment ×	0.24*	0.13	0.10	0.11
After × 7.5km	(1.68)	(0.89)	(0.68)	(0.75)
Treatment ×	0.02	-0.09*	-0.12**	-0.11**
After × 8km	(0.40)	(-1.69)	(-2.30)	(-2.04)
Treatment ×	0.14***	0.03	0.01	0.02
After × 8.5km	(2.89)	(0.43)	(0.15)	(0.31)
Treatment ×	0.68	0.57	0.53	0.55
After × 9km	(1.59)	(1.28)	(1.26)	(1.24)
Treatment ×	0.11**	-0.00	-0.03	-0.01
After × 9.5km	(2.29)	(-0.04)	(-0.54)	(-0.23)
Treatment ×	0.11**	-	-0.01	-
After × 10km	(2.11)	-	(-0.23)	-
Observations	47840	47840	47840	47840
Treatment × Ring Dummies	Yes	No	Yes	No
Year × Ring FE	Yes	Yes	Yes	Yes
Year × City FE	No	Yes	No	Yes
City × Ring FE	Yes	Yes	Yes	Yes
Dependent Variable Mean	9.62	9.62	5.86	5.86
Adj. R^2	0.375	0.563	0.243	0.475

Notes: This table reports the estimated effects of university expansion on industry patents at different distances (rings) for up to 10 km. The city-ring-specific pre-expansion time trend is removed for the dependent variables in Columns (3) and (4). The number of university teachers in 1990 is counted in 1,000, and it is used as the measure of treatment intensity. t statistics based on clustered standard errors at the city level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.9: Heterogeneity Analysis — Eastern, Central, and Western Regions

	Pre-expansion Time Trend Not Removed			Pre-expansion Time Trend Removed		
Dependent Variable	Number of Patents					
	(1)	(2)	(3)	(4)	(5)	(6)
	Eastern	Central	Western	Eastern	Central	Western
Treatment ×	114.89***	31.21***	26.51***	98.54***	25.05***	21.32***
After × 0.5km	(8.20)	(7.42)	(5.92)	(7.03)	(5.96)	(4.76)
Treatment ×	23.53***	5.74***	6.38***	20.10***	3.61**	4.39***
After × 1km	(4.21)	(3.56)	(5.19)	(3.60)	(2.24)	(3.57)
Treatment ×	7.92**	2.19***	2.46***	7.00*	1.05***	1.36**
After × 1.5km	(2.09)	(7.28)	(3.80)	(1.85)	(3.49)	(2.10)
Treatment ×	3.09*	1.88***	2.15*	2.61	1.37***	1.79
After × 2km	(1.77)	(8.17)	(2.02)	(1.50)	(5.94)	(1.68)
Treatment ×	1.18	0.84***	0.50***	0.96	0.47***	0.23*
After × 2.5km	(1.07)	(7.77)	(4.10)	(0.88)	(4.37)	(1.87)
Treatment ×	0.91	0.36***	0.71***	0.77	0.17**	0.56***
After × 3km	(1.51)	(5.25)	(6.17)	(1.29)	(2.58)	(4.84)
Treatment ×	0.43	0.42***	0.51*	0.33	0.31***	0.37
After × 3.5km	(0.71)	(5.59)	(1.83)	(0.55)	(4.09)	(1.33)
Treatment ×	-0.24	0.18**	0.27	-0.30	0.18**	0.09
After × 4km	(-1.05)	(2.63)	(1.52)	(-1.29)	(2.65)	(0.53)
Treatment ×	-0.32	0.02	0.08	-0.33	0.02	0.03
After × 4.5km	(-1.26)	(0.27)	(0.96)	(-1.29)	(0.19)	(0.33)
Treatment ×	-	-	-	-	-	-
After × 5km	-	-	-	-	-	-
Observations	10920	8320	4550	10920	8320	4550
Year × Ring FE	Yes	Yes	Yes	Yes	Yes	Yes
Year × City FE	Yes	Yes	Yes	Yes	Yes	Yes
City × Ring FE	Yes	Yes	Yes	Yes	Yes	Yes
Dependent Variable Mean	30.55	6.41	10.69	20.41	3.071	6.089
Adj. R^2	0.589	0.694	0.674	0.493	0.531	0.557

Notes: This table reports the estimated effects of university expansion on industry patents across different regions in China. The Eastern, Central and Western regions are divided according to the 7th "Five-Year Plan for the National Economic and Social Development" of China. The city-ring-specific pre-expansion time trend is removed for the dependent variables in Columns (4)-(6). The number of university teachers in 1990 is counted in 1,000, and it is used as the measure of treatment intensity. t statistics based on clustered standard errors at the city level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.10: Heterogeneity Analysis — Industries with High, Medium, and Low Human Capital Intensity

	Pre-expansion Time Trend Not Removed			Pre-expansion Time Trend Removed		
Dependent Variable	Number of Patents					
	(1)	(2)	(3)	(4)	(5)	(6)
	High	Medium	Low	High	Medium	Low
Treatment × After × 0.5km	68.58***	14.61***	3.81***	58.27***	11.65***	2.85***
	(3.72)	(5.51)	(4.73)	(3.16)	(4.39)	(3.54)
Treatment × After × 1km	14.14***	3.30***	1.01***	11.86***	2.51**	0.76***
	(4.24)	(3.03)	(3.83)	(3.55)	(2.30)	(2.88)
Treatment × After × 1.5km	4.62**	1.11**	0.50*	3.91*	0.76*	0.40
	(2.19)	(2.54)	(1.96)	(1.86)	(1.74)	(1.57)
Treatment × After × 2km	1.61**	0.77*	0.26***	1.24*	0.61	0.22***
	(2.30)	(1.68)	(3.44)	(1.77)	(1.33)	(2.97)
Treatment × After × 2.5km	0.44	0.21	0.13*	0.24	0.10	0.09
	(0.99)	(1.01)	(1.82)	(0.54)	(0.49)	(1.34)
Treatment × After × 3km	0.46	0.17	0.10	0.34	0.13	0.08
	(1.29)	(1.28)	(1.49)	(0.94)	(0.97)	(1.16)
Treatment × After × 3.5km	0.09	0.08	0.10*	0.02	0.05	0.10*
	(0.28)	(0.78)	(1.85)	(0.05)	(0.47)	(1.72)
Treatment × After × 4km	-0.32	0.06	0.03*	-0.37*	0.04	0.02
	(-1.46)	(0.46)	(1.84)	(-1.73)	(0.34)	(1.42)
Treatment × After × 4.5km	-0.34	-0.05	0.05	-0.35	-0.05	0.05
	(-1.48)	(-0.64)	(1.24)	(-1.54)	(-0.70)	(1.28)
Treatment × After × 5km	-	-	-	-	-	-
	-	-	-	-	-	-
Observations	23660	23660	23660	23660	23660	23660
Year × Ring FE	Yes	Yes	Yes	Yes	Yes	Yes
Year × City FE	Yes	Yes	Yes	Yes	Yes	Yes
City × Ring FE	Yes	Yes	Yes	Yes	Yes	Yes
Dependent Variable Mean	14.13	3.709	1.191	8.580	2.056	0.643
Adj. R^2	0.470	0.732	0.693	0.402	0.614	0.544

Notes: This table reports the estimated effects of university expansion on industry patents across industries with different human capital intensity. We define high human capital intensity industry as the industries that rank among the top one-third in the college employee ratio, medium as the middle one-third, and low as the rest. The industry college employee ratio is calculated as the percentage of workers with a college education and above using the 2004 ASIF. The city-ring-specific pre-expansion time trend is removed for the dependent variables in Columns (4)-(6). The number of university teachers in 1990 is counted in 1,000, and it is used as the measure of treatment intensity. t statistics based on clustered standard errors at the city level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.11: Effects of University Expansion on Patent Citations — Ring-level Regressions

	No. of Teachers in 1990 as Treatment		No. of Students in 1990 as Treatment	
Dependent Variable: No. of	(1) Citations to University Patents	(2) Citations to Patents beyond 5km	(3) Citations to University Patents	(4) Citations to Patents beyond 5km
Treatment ×	5.59e-01***	7.29e-02	1.14e-01**	1.52e-02
After × 0.5km	(3.24)	(1.15)	(2.59)	(0.98)
Treatment ×	8.23e-02**	4.81e-03	1.85e-02**	1.03e-03
After × 1km	(2.51)	(0.34)	(2.49)	(0.30)
Treatment ×	1.86e-02*	2.72e-02*	4.49e-03**	5.38e-03
After × 1.5km	(1.99)	(1.91)	(2.20)	(1.55)
Treatment ×	3.04e-02***	-1.41e-03	6.65e-03***	2.52e-04
After × 2km	(4.96)	(-0.10)	(4.13)	(0.08)
Treatment ×	7.07e-03	8.19e-03	1.87e-03	2.13e-03
After × 2.5km	(1.23)	(0.96)	(1.44)	(1.07)
Treatment ×	7.71e-03	1.64e-02	1.82e-03	3.88e-03*
After × 3km	(1.45)	(1.65)	(1.59)	(1.71)
Treatment ×	8.03e-03*	2.43e-02***	1.86e-03*	5.19e-03**
After × 3.5km	(1.81)	(3.14)	(1.71)	(2.57)
Treatment ×	-1.58e-03	7.77e-03	-2.62e-04	1.81e-03
After × 4km	(-0.40)	(1.16)	(-0.27)	(1.19)
Treatment ×	3.86e-03	8.22e-03	9.78e-04	1.70e-03
After × 4.5km	(1.07)	(1.28)	(1.11)	(1.08)
Treatment ×	-	-	-	-
After × 5km	-	-	-	-
Observations	4500	4500	4500	4500
Year × Ring FE	Yes	Yes	Yes	Yes
Year × City FE	Yes	Yes	Yes	Yes
City × Ring FE	Yes	Yes	Yes	Yes
Dependent Variable Mean	0.32	0.46	0.32	0.46
Adj. R^2	0.668	0.522	0.653	0.522

Notes: This table reports the estimates of the effects of university expansion on patent citations at different distances (rings). The dependent variable for Columns (1) and (3) is the ring-specific number of times when industry patents cite university patents. The dependent variable for Columns (2) and (4) is the ring-specific number of times when industry patents cite patents beyond 5 km distance from universities. The number of university teachers (students) in 1990 is used as the measure of treatment intensity in the left (right) panel. Both variables are counted in 1,000. t statistics based on clustered standard errors at the city level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.12: Effects of University Expansion on New Product Ratio — Ring-level Regressions

Dependent Variable	New Product Ratio			
	(1)	(2)	(3)	(4)
After Dummy	2000	2002	2004	2006
Treatment ×	4.16e-03**	4.84e-03**	6.35e-03**	6.68e-03**
After × 0.5km	(2.17)	(2.15)	(2.19)	(2.17)
Treatment ×	2.74e-03**	3.32e-03**	4.26e-03***	4.58e-03**
After × 1km	(2.57)	(2.59)	(2.68)	(2.46)
Treatment ×	1.76e-03***	2.07e-03***	2.74e-03***	2.89e-03***
After × 1.5km	(4.98)	(5.17)	(4.87)	(4.60)
Treatment ×	1.28e-03***	1.47e-03***	1.75e-03***	1.77e-03***
After × 2km	(4.91)	(4.80)	(5.14)	(4.78)
Treatment ×	1.66e-03***	2.03e-03***	1.73e-03***	1.63e-03***
After × 2.5km	(3.65)	(3.90)	(3.62)	(3.20)
Treatment ×	1.07e-03***	1.31e-03***	1.63e-03***	1.68e-03***
After × 3km	(4.51)	(5.10)	(4.69)	(4.59)
Treatment ×	6.20e-04***	7.28e-04**	1.04e-03***	1.21e-03***
After × 3.5km	(2.74)	(2.54)	(3.12)	(3.39)
Treatment ×	6.66e-04***	7.85e-04***	1.24e-03***	1.56e-03***
After × 4km	(2.78)	(2.99)	(3.97)	(4.31)
Treatment ×	6.06e-04	6.96e-04	9.91e-04	1.14e-03**
After × 4.5km	(1.19)	(1.24)	(1.61)	(2.01)
Treatment ×	5.54e-04**	6.84e-04**	1.11e-03***	1.38e-03***
After × 5km	(2.17)	(2.13)	(2.84)	(3.38)
Observations	1196263	996185	759980	589233
Year × Ring FE	Yes	Yes	Yes	Yes
Year × City FE	Yes	Yes	Yes	Yes
City × Ring FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Control Variables	Yes	Yes	Yes	Yes
Dependent Variable Mean	0.034	0.035	0.037	0.037
Adj. R^2	0.091	0.097	0.107	0.106

Notes: This table reports the estimated effects of university expansion on firms' new product ratio using the number of university teachers in 1990 as the proxy for treatment intensity. The dependent variable is firm-level new product ratio. Columns (1)–(4) report the triple-differences estimates. The after dummy equals 1 if year is 2000 or after, 2002 or after, 2004 or after, or 2006 or after in Columns (1), (2), (3), and (4), respectively. The after dummy equals 0 if year is before 2000 for all four columns. Observations in the years in which the after dummy is not defined are dropped. The reference group is the firms outside 10 km of universities. Control variables include firm age, fixed assets, SOE status, and employment size. The number of university teachers in 1990 is counted in 1,000. t statistics based on clustered standard errors at the city level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.13: Heterogeneity Analysis — Industries with High, Medium, and Low Human Capital Intensity and SOE versus Non-SOE

Dependent Variable	New Product Ratio				
	(1) High	(2) Medium	(3) Low	(4) SOE	(5) Non-SOE
Treatment × After × 0.5km	4.70e-03**	2.77e-03**	1.41e-03	8.92e-04	4.31e-03**
	(2.20)	(2.43)	(1.13)	(1.15)	(1.99)
Treatment × After × 1km	3.88e-03**	1.41e-03**	1.08e-03**	9.88e-04*	2.99e-03***
	(2.51)	(2.10)	(2.53)	(1.92)	(2.62)
Treatment × After × 1.5km	2.79e-03***	1.29e-03***	1.48e-04	7.21e-04***	1.91e-03***
	(4.20)	(5.00)	(0.80)	(2.89)	(4.91)
Treatment × After × 2km	1.52e-03**	8.51e-04***	7.62e-04***	5.37e-04*	1.44e-03***
	(2.43)	(3.29)	(3.79)	(1.70)	(6.24)
Treatment × After × 2.5km	2.99e-03**	5.98e-04**	6.66e-04***	1.36e-03***	1.74e-03***
	(2.17)	(2.30)	(3.16)	(4.99)	(3.52)
Treatment × After × 3km	1.54e-03***	8.52e-04**	2.58e-04	8.97e-04**	1.21e-03***
	(3.73)	(2.03)	(1.29)	(2.08)	(5.76)
Treatment × After × 3.5km	7.64e-04*	4.43e-04	2.49e-04	3.63e-05	8.57e-04***
	(1.78)	(1.14)	(0.80)	(0.08)	(3.41)
Treatment × After × 4km	7.00e-04	6.94e-05	6.37e-04**	-1.96e-04	9.36e-04***
	(0.67)	(0.24)	(2.02)	(-0.41)	(3.16)
Treatment × After × 4.5km	8.83e-04*	-2.00e-04	9.10e-04***	-6.10e-04*	9.00e-04*
	(1.67)	(-0.34)	(3.78)	(-1.78)	(1.89)
Treatment × After × 5km	-2.63e-04	5.37e-04	7.76e-04***	1.67e-04	8.22e-04***
	(-0.46)	(0.93)	(3.12)	(0.33)	(3.42)
Observations	394427	385023	456632	136171	1060023
Year × Ring FE	Yes	Yes	Yes	Yes	Yes
Year × City FE	Yes	Yes	Yes	Yes	Yes
City × Ring FE	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes
Control Variables	Yes	Yes	Yes	Yes	Yes
Dependent Variable Mean	0.059	0.025	0.020	0.046	0.033
Adj. R^2	0.119	0.057	0.049	0.111	0.096

Notes: Columns (1)–(3) report the estimated effects of university expansion on firms' new product ratio across industries with different human capital intensity. We define high human capital intensity industry as the industries that rank among the top one-third in the college employee ratio, medium as the middle one-third, and low as the rest. The industry college employee ratio is calculated as the percentage of workers with a college education and above using the 2004 ASIF. Columns (4) and (5) report the estimates of the effects of university expansion on firms' new product ratio for SOEs and non-SOEs separately. The number of university teachers in 1990 is counted in 1,000, and it is used as the treatment intensity. Control variables include firm age, fixed assets, SOE status, and employment size. The reference group consists of the firms outside 10 km of universities. t statistics based on clustered standard errors at the city level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix

Appendix A: Figures and Tables

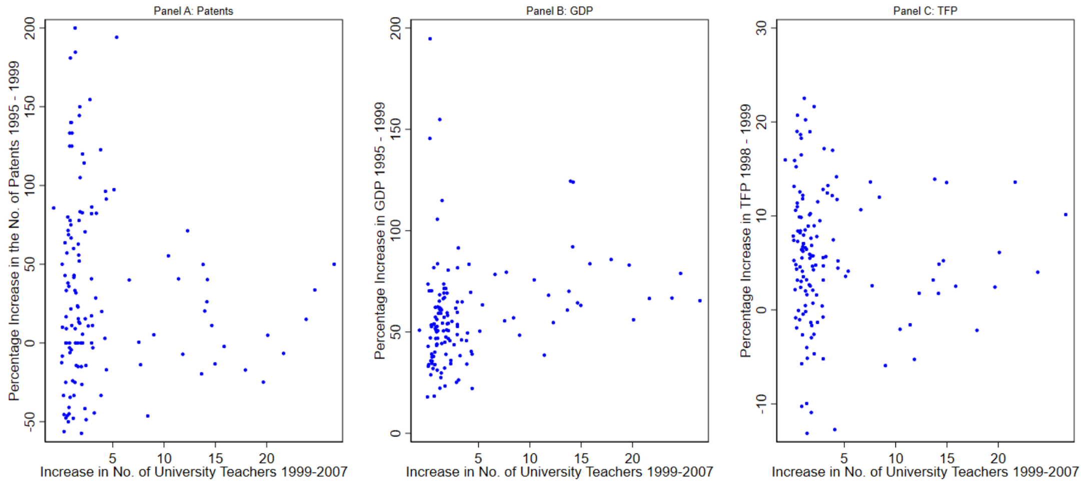


Figure A1: The Extent of University Expansion and Pre-expansion Growth of Patents, GDP and TFP

Notes: Panel A shows the scatter plot of pre-expansion growth of patents against the extent of university expansion at the city level. Panel B shows the scatter plot of pre-expansion growth of GDP against the extent of university expansion at the city level. Panel C shows the scatter plot of pre-expansion growth of average firm TFP against the extent of university expansion at the city level. The correlation coefficients are -0.10, 0.24 and -0.05 respectively. The number of university teachers is counted in 1,000.

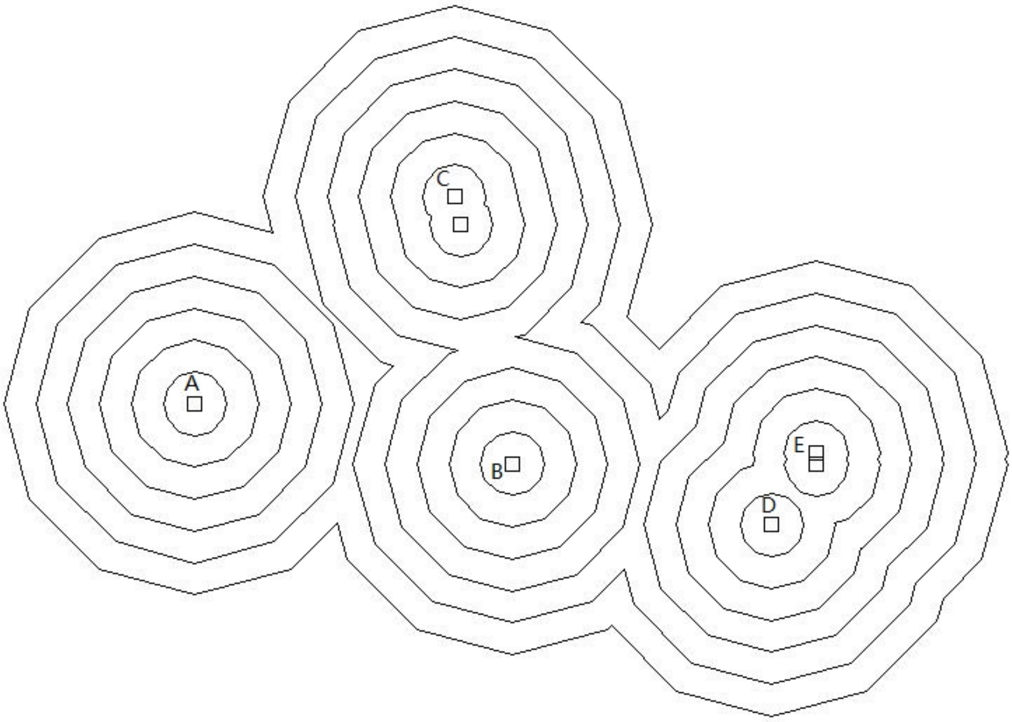
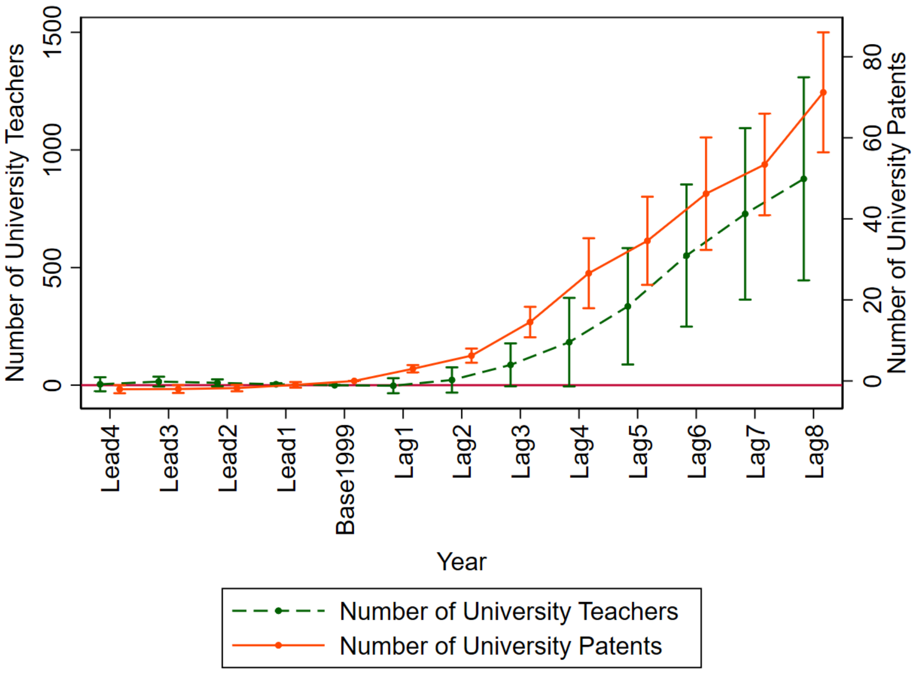
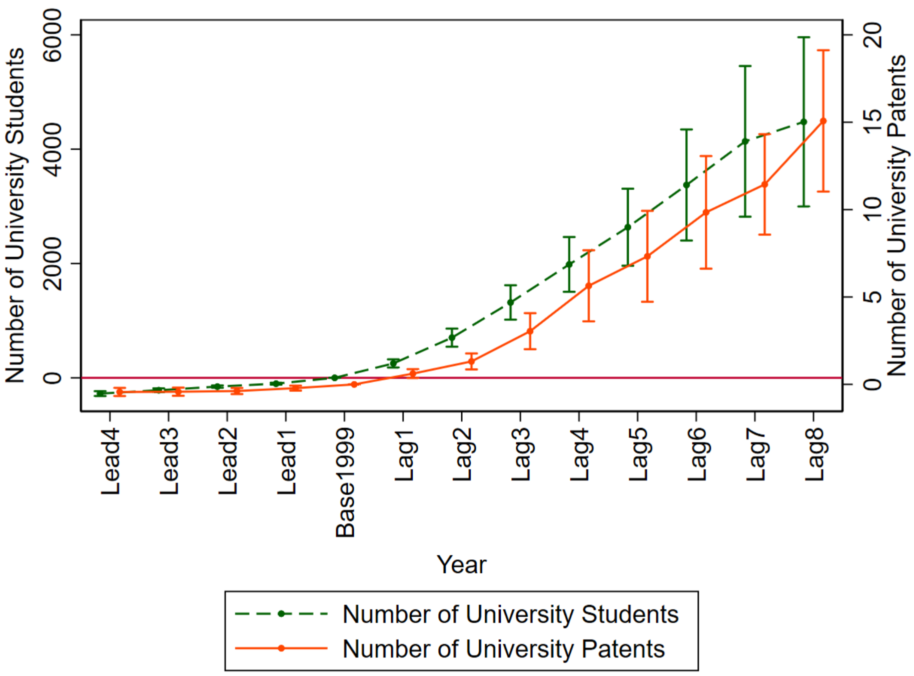


Figure A2: Illustrative Graph for the Construction of Concentric Rings

Notes: The centers of the rings comprise the locations of universities and entities that have direct collaborations with universities. The two locations at C belong to university I. The three locations at D and E belong to university II. One university can have multiple locations in the dataset because the address filed in a patent application points to the exact building of the patent applicant. Point A stands for a non-university entity that has direct collaboration with university I. Point B stands for a non-university entity that has direct collaboration with university II.

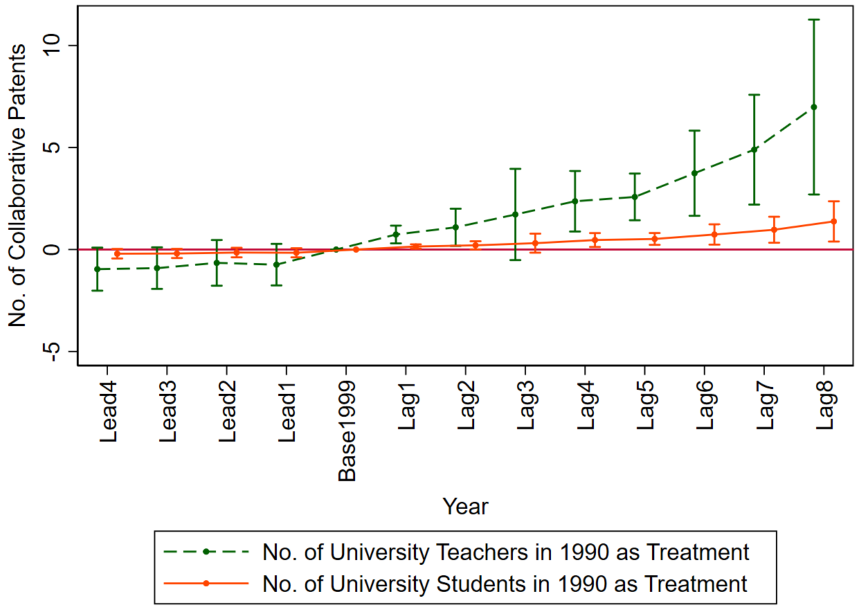


(a) No. of University Teachers in 1990 as Treatment

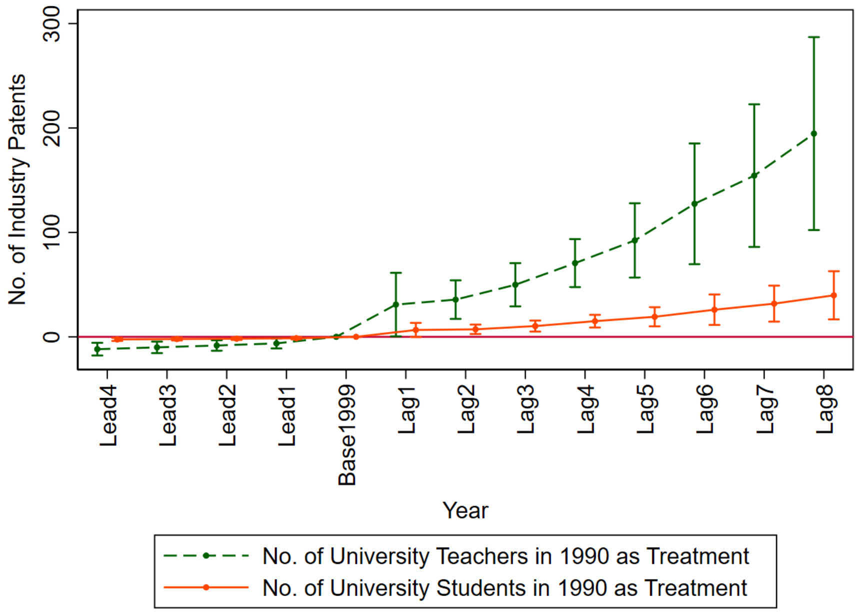


(b) No. of University Students in 1990 as Treatment

Figure A3: The Dynamic Effects of University Expansion on the Numbers of University Teachers, Students, and University Patents – Pre-expansion Time Trend Not Removed

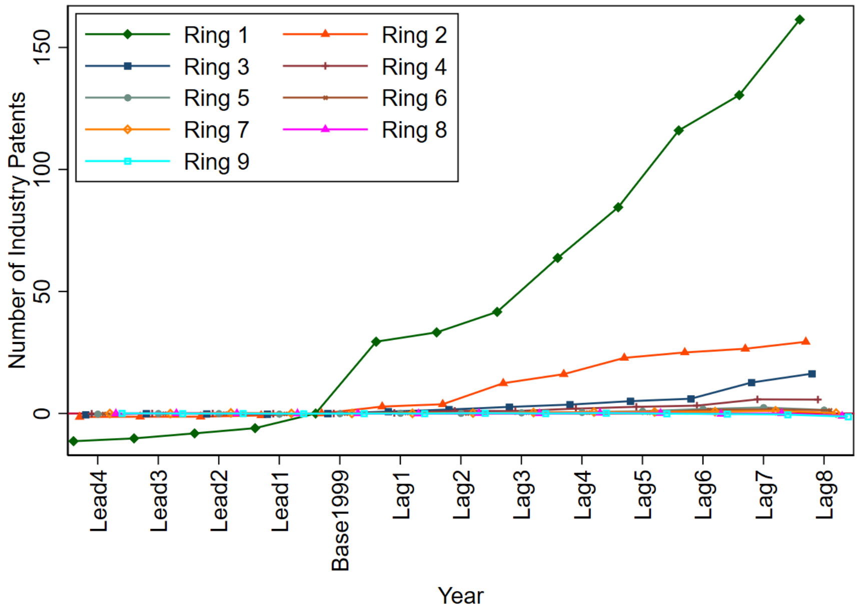


(a) Collaborative Patents

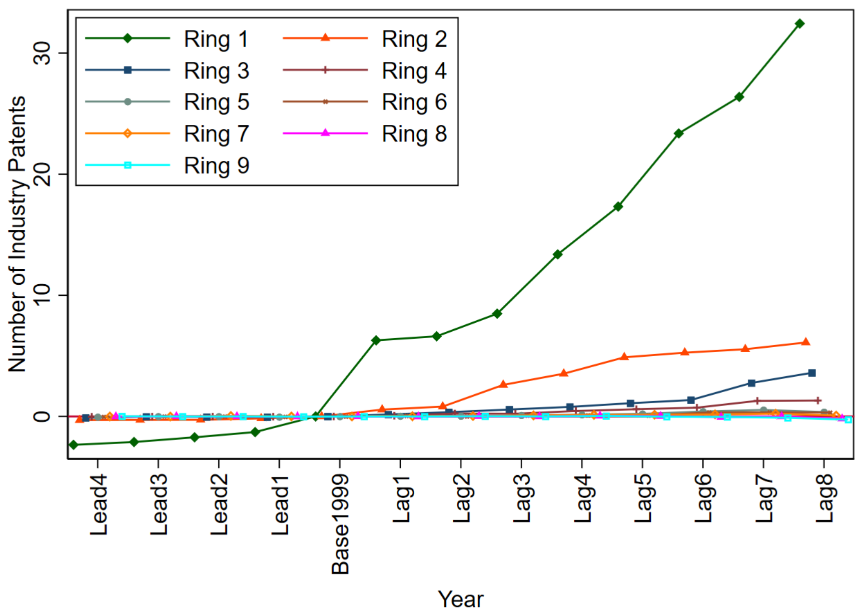


(b) Industry Patents

Figure A4: The Dynamic Effects of University Expansion on the Numbers of Collaborative Patents and Industry Patents – Pre-expansion Time Trend Not Removed



(a) No. of University Teachers in 1990 as Treatment



(b) No. of University Students in 1990 as Treatment

Figure A5: The Dynamic Effects of University Expansion on the Number of Industry Patents at the Ring Level — Pre-expansion Time Trend Not Removed

Table A1.A: Impact of University Expansion on University Scale and and Innovation

	Pre-expansion Time Trend Not Removed					Pre-expansion Time Trend Removed				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dependent Variable: No. of	University Students	Total Patents	University Patents	Collabo-rative Patents	Industry Patents	University Students	Total Patents	University Patents	Collabo-rative Patents	Industry Patents
Treatment × After	2509.57***	28.83***	7.08***	0.73***	21.02***	2059.45***	24.31***	6.46***	0.48**	17.37***
	(7.25)	(4.20)	(6.53)	(3.20)	(3.70)	(5.95)	(3.54)	(5.96)	(2.11)	(3.06)
Observations	2352	2352	2352	2352	2352	2352	2352	2352	2352	2352
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control Variables	No	No	No	No	No	No	No	No	No	No
Dependent Variable Mean	39387.99	222.44	31.38	3.38	187.68	20274.18	136.63	23.86	1.13	111.64
Adj. R^2	0.829	0.575	0.638	0.652	0.530	0.657	0.466	0.579	0.469	0.419

Notes: This table reports the estimated effects of university expansion on the numbers of university students and different classifications of patents. The number of university students in 1990 is counted in 1,000, and it is used as the measure of treatment intensity. The number of university students is considered as a proxy for university scale. t statistics based on clustered standard errors at the city level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A1.B: Impact of University Expansion on University Scale and and Innovation — Robustness

	Pre-expansion Time Trend Not Removed					Pre-expansion Time Trend Removed				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dependent Variable: No. of	University Teachers	Total Patents	University Patents	Collabo-rative Patents	Industry Patents	University Teachers	Total Patents	University Patents	Collabo-rative Patents	Industry Patents
Treatment × After	220.99**	113.88***	28.83***	3.31***	81.75***	232.85**	93.35***	25.86***	2.11**	65.38***
	(2.28)	(4.22)	(7.68)	(3.66)	(3.44)	(2.40)	(3.46)	(6.90)	(2.34)	(2.75)
Observations	2330	2338	2338	2338	2338	2330	2338	2338	2338	2338
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dependent Variable Mean	3032.32	225.23	31.93	3.43	189.87	802.97	141.56	25.38	1.66	114.52
Adj. R^2	0.930	0.604	0.674	0.686	0.556	0.759	0.497	0.620	0.500	0.443

Notes: This table reports the estimated effects of university expansion on the numbers of university teachers and different classifications of patents. The number of university teachers in 1990 is counted in 1,000, and it is used as the measure of treatment intensity. The number of university teachers is considered as a proxy for university scale. Control variables include the non-agricultural population, the proportion of employment in the manufacturing industries, and the proportion of employment in the service industries. t statistics based on clustered standard errors at the city level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A1.C: Impact of University Expansion on University Scale and and Innovation — Robustness

	Pre-expansion Time Trend Not Removed					Pre-expansion Time Trend Removed				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dependent Variable: No. of	University Students	Total Patents	University Patents	Collabo-rative Patents	Industry Patents	University Students	Total Patents	University Patents	Collabo-rative Patents	Industry Patents
Treatment × After	1991.77***	23.06***	6.05***	0.64***	16.37***	1602.58***	18.87***	5.44***	0.39*	13.04**
	(5.73)	(3.44)	(6.35)	(2.97)	(2.86)	(4.73)	(2.81)	(5.72)	(1.82)	(2.28)
Observations	2338	2338	2338	2338	2338	2338	2338	2338	2338	2338
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dependent Variable Mean	39659.82	225.23	31.93	3.43	189.87	19914.02	140.17	25.17	1.58	113.42
Adj. R^2	0.860	0.595	0.663	0.664	0.548	0.714	0.488	0.609	0.478	0.437

Notes: This table reports the estimated effects of university expansion on the numbers of university students and different classifications of patents. The number of university students in 1990 is counted in 1,000, and it is used as the measure of treatment intensity. The number of university students is considered as a proxy for university scale. Control variables include the non-agricultural population, the proportion of employment in the manufacturing industries, and the proportion of employment in the service industries. t statistics based on clustered standard errors at the city level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A2: Impact of University Expansion on University Scale and Innovation — City-level Analysis of Trend Break Model

	(1)	(2)	(3)	(4)	(5)
Dependent Variable: No. of	University Students	Total Patents	University Patents	Collabo-rative Patents	Industry Patents
Treatment × Trend	609.1***	6.690***	2.153***	0.126**	4.411***
×After 2000	(5.28)	(4.33)	(8.01)	(2.20)	(3.41)
Observations	2392	2392	2392	2392	2392
Year FE	Yes	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes	Yes
Dependent Variable Mean	3006.14	222.44	31.38	3.38	187.68
Adj. R^2	0.924	0.660	0.854	0.742	0.586

Notes: This table reports the estimates of the slope change in the numbers of university students and different classifications patents as a result of the university expansion, using the specification in Equation (3.7). The number of university students in 1990 is counted in 1,000, and it is used as the measure of treatment intensity. t statistics based on clustered standard errors at the city level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A3.A: The Dynamic Effects of University Expansion on the Number of Teachers and Innovation

	Pre-expansion Time Trend Not Removed					Pre-expansion Time Trend Removed				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dependent Variable: No. of	University Teachers	Total Patents	University Patents	Collabo-rative Patents	Industry Patents	University Teachers	Total Patents	University Patents	Collabo-rative Patents	Industry Patents
Treatment × 1995	3.97	-14.83***	-2.04***	-0.97*	-11.83***	1.01	-1.17	-0.16	-0.21	-0.81
	(0.26)	(-3.73)	(-4.03)	(-1.80)	(-3.84)	(0.07)	(-0.30)	(-0.31)	(-0.39)	(-0.26)
Treatment × 1996	14.90	-12.98***	-1.99***	-0.91*	-10.08***	12.68	-2.74	-0.58	-0.35	-1.81
	(1.41)	(-3.49)	(-4.03)	(-1.77)	(-3.61)	(1.20)	(-0.74)	(-1.17)	(-0.67)	(-0.65)
Treatment × 1997	10.22	-10.71***	-1.75***	-0.66	-8.30***	8.74	-3.89	-0.81**	-0.28	-2.79
	(1.37)	(-3.14)	(-4.35)	(-1.16)	(-3.25)	(1.17)	(-1.14)	(-2.02)	(-0.49)	(-1.09)
Treatment × 1998	4.04	-7.99**	-0.96***	-0.74	-6.28**	3.30	-4.57	-0.49	-0.55	-3.53
	(1.63)	(-2.42)	(-2.71)	(-1.44)	(-2.54)	(1.33)	(-1.39)	(-1.39)	(-1.07)	(-1.43)
Treatment × 2000	-2.19	34.70**	3.02***	0.74***	30.95**	-1.45	31.29**	2.54***	0.55**	28.20*
	(-0.13)	(2.21)	(6.58)	(3.34)	(2.00)	(-0.09)	(1.99)	(5.55)	(2.49)	(1.83)
Treatment × 2001	21.81	43.06***	6.27***	1.09**	35.71***	23.29	36.24***	5.32***	0.71	30.20***
	(0.80)	(4.19)	(7.08)	(2.36)	(3.81)	(0.86)	(3.53)	(6.01)	(1.55)	(3.22)
Treatment × 2002	85.72*	66.17***	14.57***	1.72	49.88***	87.94*	55.92***	13.15***	1.15	41.61***
	(1.86)	(5.26)	(7.66)	(1.52)	(4.74)	(1.90)	(4.45)	(6.92)	(1.02)	(3.96)
Treatment × 2003	181.77*	99.63***	26.63***	2.37***	70.63***	184.72*	85.98***	24.75***	1.61**	59.61***
	(1.91)	(6.80)	(6.08)	(3.15)	(6.05)	(1.94)	(5.87)	(5.65)	(2.15)	(5.11)
Treatment × 2004	333.73***	129.62***	34.66***	2.58***	92.37***	337.43***	112.54***	32.31***	1.64***	78.60***
	(2.67)	(5.76)	(6.28)	(4.46)	(5.11)	(2.70)	(5.00)	(5.86)	(2.83)	(4.35)
Treatment × 2005	549.53***	177.42***	46.31***	3.74***	127.36***	553.96***	156.93***	43.49***	2.61**	110.83***
	(3.59)	(5.16)	(6.58)	(3.54)	(4.34)	(3.62)	(4.57)	(6.18)	(2.47)	(3.77)
Treatment × 2006	726.80***	212.83***	53.51***	4.90***	154.42***	731.97***	188.93***	50.22***	3.58***	135.14***
	(3.93)	(5.38)	(8.41)	(3.59)	(4.45)	(3.96)	(4.78)	(7.89)	(2.62)	(3.90)
Treatment × 2007	875.96***	273.44***	71.41***	7.00***	195.04***	881.87***	246.13***	67.64***	5.49**	173.00***
	(4.00)	(5.06)	(9.52)	(3.22)	(4.16)	(4.03)	(4.55)	(9.02)	(2.52)	(3.69)
Observations	2344	2352	2352	2352	2352	2344	2352	2352	2352	2352
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dependent Variable Mean	3016.79	223.97	31.74	3.41	188.82	832.77	139.57	24.41	1.26	113.91
Adj. R^2	0.954	0.682	0.885	0.794	0.604	0.847	0.580	0.856	0.624	0.491

Notes: This table reports the estimates of the dynamic effects of university expansion on the number of university teachers and and different classifications of patents. The estimates are used to plot Figure 3, Figure 4, Appendix Figure A3, and Appendix Figure A4. The number of university teachers in 1990 is used as the treatment intensity measure, and it is counted in 1,000. The base year is 1999. t statistics based on clustered standard errors at the city level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A3.B: The Dynamic Effects of University Expansion on the Number of Students and Innovation

	Pre-expansion Time Trend Not Removed					Pre-expansion Time Trend Removed				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dependent Variable: No. of	University Students	Total Patents	University Patents	Collabo-rative Patents	Industry Patents	University Students	Total Patents	University Patents	Collabo-rative Patents	Industry Patents
Treatment × 1995	-275.17***	-3.11***	-0.43***	-0.21*	-2.47***	1.64	-0.32	-0.05	-0.05	-0.22
	(-12.72)	(-3.23)	(-3.67)	(-1.73)	(-3.28)	(0.08)	(-0.33)	(-0.43)	(-0.43)	(-0.29)
Treatment × 1996	-215.85***	-2.74***	-0.42***	-0.19*	-2.12***	-8.25	-0.65	-0.13	-0.08	-0.43
	(-13.21)	(-3.11)	(-3.56)	(-1.71)	(-3.18)	(-0.50)	(-0.73)	(-1.13)	(-0.69)	(-0.65)
Treatment × 1997	-152.82***	-2.30***	-0.38***	-0.15	-1.77***	-14.41	-0.90	-0.19**	-0.07	-0.64
	(-15.09)	(-2.95)	(-4.29)	(-1.25)	(-2.97)	(-1.42)	(-1.16)	(-2.13)	(-0.60)	(-1.08)
Treatment × 1998	-101.87***	-1.72**	-0.21***	-0.16	-1.35**	-32.67***	-1.02	-0.12	-0.12	-0.79
	(-12.52)	(-2.37)	(-2.98)	(-1.40)	(-2.44)	(-4.02)	(-1.41)	(-1.65)	(-1.06)	(-1.42)
Treatment × 2000	251.56***	7.36**	0.62***	0.15***	6.60*	182.36***	6.67*	0.52***	0.11**	6.04*
	(6.94)	(2.10)	(4.81)	(2.75)	(1.93)	(5.03)	(1.90)	(4.07)	(2.03)	(1.76)
Treatment × 2001	702.15***	8.74***	1.32***	0.21**	7.22***	563.75***	7.35***	1.13***	0.13	6.09***
	(8.77)	(3.41)	(5.68)	(2.03)	(3.15)	(7.04)	(2.87)	(4.86)	(1.27)	(2.66)
Treatment × 2002	1316.18***	13.69***	3.05***	0.31	10.33***	1108.57***	11.60***	2.76***	0.20	8.64***
	(8.66)	(4.18)	(5.84)	(1.32)	(3.84)	(7.29)	(3.54)	(5.29)	(0.83)	(3.22)
Treatment × 2003	1981.72***	21.16***	5.65***	0.46***	15.04***	1704.92***	18.37***	5.27***	0.31*	12.79***
	(8.16)	(5.40)	(5.48)	(2.69)	(4.88)	(7.02)	(4.69)	(5.11)	(1.80)	(4.15)
Treatment × 2004	2618.09***	27.08***	7.35***	0.52***	19.21***	2272.09***	23.59***	6.87***	0.33**	16.40***
	(7.76)	(4.60)	(5.57)	(3.53)	(4.13)	(6.74)	(4.01)	(5.21)	(2.22)	(3.52)
Treatment × 2005	3363.46***	36.56***	9.87***	0.74***	25.95***	2948.25***	32.38***	9.29***	0.51**	22.58***
	(6.85)	(4.09)	(6.04)	(2.92)	(3.50)	(6.00)	(3.62)	(5.69)	(2.00)	(3.05)
Treatment × 2006	4123.38***	44.23***	11.47***	0.97***	31.79***	3638.98***	39.35***	10.80***	0.70**	27.85***
	(6.19)	(4.32)	(7.88)	(3.00)	(3.63)	(5.46)	(3.84)	(7.42)	(2.17)	(3.18)
Treatment × 2007	4465.99***	56.34***	15.12***	1.38***	39.84***	3912.38***	50.77***	14.35***	1.07**	35.34***
	(5.96)	(4.03)	(7.38)	(2.75)	(3.40)	(5.22)	(3.63)	(7.01)	(2.14)	(3.02)
Observations	2352	2352	2352	2352	2352	2352	2352	2352	2352	2352
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dependent Variable Mean	39448.11	223.97	31.74	3.41	188.82	20221.56	137.52	24.14	1.14	112.24
Adj. R^2	0.926	0.662	0.865	0.750	0.587	0.830	0.558	0.834	0.571	0.474

Notes: This table reports the estimates of the dynamic effects of university expansion on the number of university students and different classifications of patents. The estimates are used to plot Figure 3, Figure 4, Appendix Figure A3, and Appendix Figure A4. The number of university students in 1990 is used as the treatment intensity measure, and it is counted in 1,000. The base year is 1999. t statistics based on clustered standard errors at the city level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A4.A: 2SLS Estimates — Effects of University Innovation Capacity on Local Innovation Activities
(Pre-expansion Time Trend Not Removed)

	(1)	(2)	(3)	(4)
	Panel A: No. of University Teachers in 1990 Interacted with Post-expansion Dummy as the IV			
	No. of University Teachers as First-Stage Dependent Variable		No. of University Patents as First-Stage Dependent Variable	
Dependent Variable: No. of	Collaborative Patents	Industry Patents	Collaborative Patents	Industry Patents
University Innovation Capacity	10.754*	298.337**	110.039***	3053.193***
	(1.81)	(1.99)	(4.40)	(6.53)
Observations	2384	2384	2392	2392
First-stage F-statistics	9.616	9.616	52.902	52.902
Dependent Variable Mean	3.389	188.201	3.378	187.677
Adj. R^2	0.106	0.255	0.829	0.639
	Panel B: No. of University Students in 1990 Interacted with Post-expansion Dummy as the IV			
	No. of University Students as First-Stage Dependent Variable		No. of University Patents as First-Stage Dependent Variable	
Dependent Variable: No. of	Collaborative Patents	Industry Patents	Collaborative Patents	Industry Patents
University Innovation Capacity	0.292**	8.375**	103.5***	2969.6***
	(2.28)	(2.51)	(4.55)	(6.63)
Observations	2392	2392	2392	2392
First-stage F-statistics	48.453	48.453	39.380	39.380
Dependent Variable Mean	3.378	187.677	3.378	187.677
Adj. R^2	0.523	0.463	0.835	0.640
Year FE	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes

Notes: This table reports the 2SLS estimates of the effects of university innovation capacity on innovation activities at the city level, using the number of university teachers or students in 1990 interacted with the after dummy as the instrument. All the First-Stage Dependent Variables are counted in 1,000. The F-statistics is calculated based on Montiel Olea and Pflueger (2013), which is robust to heteroskedasticity, autocorrelation, and clustering. t statistics based on clustered standard errors at the city level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A4.B: 2SLS Estimates — Effects of University Innovation Capacity on Local Innovation Activities (Pre-expansion Time Trend Removed)

	(1)	(2)	(3)	(4)
	Panel A: No. of University Teachers in 1990 Interacted with Post-expansion Dummy as the IV			
	No. of University Teachers as First-Stage Dependent Variable		No. of University Patents as First-Stage Dependent Variable	
Dependent Variable: No. of	Collaborative Patents	Industry Patents	Collaborative Patents	Industry Patents
University Innovation Capacity	7.062	242.401*	80.706***	2771.068***
	(1.49)	(1.85)	(2.94)	(5.29)
Observations	2384	2384	2392	2392
First-stage F-statistics	9.903	9.903	43.663	43.663
Dependent Variable Mean	1.246	113.626	1.242	113.270
Adj. R^2	0.005	0.181	0.711	0.539
	Panel B: No. of University Students in 1990 Interacted with Post-expansion Dummy as the IV			
	No. of University Students as First-Stage Dependent Variable		No. of University Patents as First-Stage Dependent Variable	
Dependent Variable: No. of	Collaborative Patents	Industry Patents	Collaborative Patents	Industry Patents
University Innovation Capacity	0.234	8.435**	74.679***	2690.238***
	(1.59)	(2.07)	(2.76)	(5.19)
Observations	2392	2392	2392	2392
First-stage F-statistics	32.631	32.631	32.787	32.787
Dependent Variable Mean	1.127	111.641	1.127	111.641
Adj. R^2	0.283	0.316	0.715	0.544
Year FE	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes

Notes: This table reports the 2SLS estimates of the effects of university innovation capacity on innovation activities at the city level, using the number of university teachers or students in 1990 interacted with the after dummy as the instrument. All the First-Stage Dependent Variables are counted in 1,000. The F-statistics is calculated based on Montiel Olea and Pflueger (2013), which is robust to heteroskedasticity, autocorrelation, and clustering. The city-specific pre-expansion time trend is removed for the dependent variable in all specifications. t statistics based on clustered standard errors at the city level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A5: Impact of University Expansion on Industry Innovation — Ring-level Regressions

	Pre-expansion Time Trend Not Removed			Pre-expansion Time Trend Removed		
Dependent Variable	Number of Patents					
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment ×	18.43***	18.28***	0.74	15.72***	15.64***	0.64
After × 0.5km	(3.61)	(3.60)		(3.08)	(3.08)	
Treatment ×	4.02***	3.87***	0.53	3.33***	3.24***	0.45
After × 1km	(3.91)	(3.89)		(3.24)	(3.26)	
Treatment ×	1.53***	1.38**	0.35	1.25**	1.17**	0.29
After × 1.5km	(2.71)	(2.58)		(2.21)	(2.18)	
Treatment ×	0.78***	0.63***	0.25	0.61**	0.52**	0.21
After × 2km	(2.93)	(2.65)		(2.28)	(2.21)	
Treatment ×	0.40**	0.25	0.13	0.28	0.19	0.10
After × 2.5km	(2.24)	(1.64)		(1.53)	(1.26)	
Treatment ×	0.33***	0.18**	0.14	0.23**	0.14*	0.12
After × 3km	(2.89)	(2.05)		(2.00)	(1.67)	
Treatment ×	0.26**	0.11	0.13	0.17	0.08	0.10
After × 3.5km	(2.42)	(1.24)		(1.57)	(0.99)	
Treatment ×	0.13***	-0.02	-0.03	0.05	-0.03	-0.05
After × 4km	(2.97)	(-0.48)		(1.16)	(-0.83)	
Treatment ×	0.11**	-0.05	-0.10	0.04	-0.05	-0.10
After × 4.5km	(2.15)	(-1.19)		(0.72)	(-1.25)	
Treatment ×	0.15***	-	-	0.08	-	-
After × 5km	(3.00)	-		(1.64)	-	
Observations	23920	23920	-	23920	23920	-
Treatment × Ring dummies	Yes	No	-	Yes	No	-
City FE	Yes	No	-	Yes	No	-
Year × Ring FE	Yes	Yes	-	Yes	Yes	-
Year × City FE	No	Yes	-	No	Yes	-
City × Ring FE	No	Yes	-	No	Yes	-
Dependent Variable Mean	18.21	18.21	-	10.59	10.59	-
Adjusted R^2	0.351	0.560	-	0.228	0.479	-

Notes: This table reports the estimates of the effects of university expansion on industry patents at different distances (rings). The city-ring-specific pre-expansion time trend is removed for the dependent variable in columns (4)-(5). Column (3) and (6) are obtained by dividing the coefficients in column (2) and (5) by the average number of patents in the corresponding ring during the pre-expansion periods. The number of university students in 1990 is counted in 1,000, and it is used as the measure of treatment intensity. t statistics based on clustered standard errors at the city level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A6.A: Impact of University Expansion on Innovation — Ring Regressions (Robustness)

	Pre-expansion Time Trend Not Removed		Pre-expansion Time Trend Removed	
Dependent Variable	Number of Patents that are Cited at least Once			
	(1)	(2)	(3)	(4)
Treatment ×	22.79***	19.69**	16.80*	16.38**
After × 0.5km	(2.67)	(2.59)	(1.97)	(2.15)
Treatment ×	4.92***	4.43***	4.12***	3.72***
After × 1km	(5.74)	(5.26)	(4.81)	(4.42)
Treatment ×	1.51**	1.56**	1.74**	1.36**
After × 1.5km	(2.10)	(2.56)	(2.41)	(2.24)
Treatment ×	0.66	0.80**	1.07**	0.69*
After × 2km	(1.24)	(2.06)	(2.02)	(1.79)
Treatment ×	0.05	0.26	0.59*	0.22
After × 2.5km	(0.16)	(1.32)	(1.75)	(1.11)
Treatment ×	-0.04	0.17	0.51*	0.13
After × 3km	(-0.14)	(1.34)	(1.94)	(1.04)
Treatment ×	-0.18	0.05	0.40*	0.03
After × 3.5km	(-0.80)	(0.46)	(1.73)	(0.27)
Treatment ×	-0.29	-0.05	0.30	-0.07
After × 4km	(-1.50)	(-0.61)	(1.58)	(-0.80)
Treatment ×	-0.34*	-0.10	0.26	-0.11
After × 4.5km	(-1.83)	(-1.31)	(1.44)	(-1.43)
Treatment ×	-0.25	-	0.37**	-
After × 5km	(-1.40)	-	(2.07)	-
Observations	8600	8600	8600	8600
Treatment × Ring Dummies	Yes	No	Yes	No
City FE	Yes	No	Yes	No
Year × Ring FE	Yes	Yes	Yes	Yes
Year × City FE	No	Yes	No	Yes
City × Ring FE	No	Yes	No	Yes
Dependent Variable Mean	7.45	7.45	5.14	5.14
Adj. R^2	0.246	0.473	0.158	0.410

Notes: This table reports the estimates of the effects of university expansion on industry patents with at least one citation at different distances (rings). The city-ring-specific pre-expansion time trend is removed for the dependent variable in columns (3)-(4). The number of university teachers in 1990 is used as the measure of treatment intensity, and it is counted in 1,000. t statistics based on clustered standard errors at the city level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A6.B: Impact of University Expansion on Innovation — Ring Regressions (Robustness)

	Pre-expansion Time Trend Not Removed		Pre-expansion Time Trend Removed	
Dependent Variable	Number of Patents that are Cited at least Once			
	(1)	(2)	(3)	(4)
Treatment ×	4.49**	3.87**	3.31	3.22*
After × 0.5km	(2.25)	(2.17)	(1.65)	(1.80)
Treatment ×	1.04***	0.93***	0.87***	0.78***
After × 1km	(4.54)	(4.17)	(3.80)	(3.51)
Treatment ×	0.35**	0.35**	0.38**	0.30**
After × 1.5km	(2.32)	(2.62)	(2.58)	(2.28)
Treatment ×	0.16	0.18**	0.24**	0.16*
After × 2km	(1.50)	(2.27)	(2.26)	(1.98)
Treatment ×	0.02	0.06	0.13*	0.05
After × 2.5km	(0.33)	(1.40)	(1.86)	(1.17)
Treatment ×	0.00	0.04	0.11**	0.03
After × 3km	(0.02)	(1.34)	(2.04)	(1.04)
Treatment ×	-0.03	0.01	0.09*	0.01
After × 3.5km	(-0.64)	(0.56)	(1.85)	(0.39)
Treatment ×	-0.05	-0.01	0.07	-0.01
After × 4km	(-1.30)	(-0.53)	(1.56)	-0.70)
Treatment ×	-0.07	-0.02	0.06	-0.02
After × 4.5km	(-1.62)	(-1.22)	(1.39)	(-1.33)
Treatment ×	-0.05	-	0.08**	-
After × 5km	(-1.22)	-	(2.05)	-
Observations	8600	8600	8600	8600
Treatment × Ring Dummies	Yes	No	Yes	No
City FE	Yes	No	Yes	No
Year × Ring FE	Yes	Yes	Yes	Yes
Year × City FE	No	Yes	No	Yes
City × Ring FE	No	Yes	No	Yes
Dependent Variable Mean	7.45	7.45	5.13	5.13
Adj. R^2	0.216	0.460	0.138	0.405

Notes: This table reports the estimates of the effects of university expansion on industry patents with at least one citation at different distances (rings). The city-ring-specific pre-expansion time trend is removed for the dependent variable in columns (3)-(4). The number of university students in 1990 is used as the measure of treatment intensity, and it is counted in 1,000. t statistics based on clustered standard errors at the city level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A7.A: The Dynamic Effects of University Expansion on Industry Innovation — Ring Regressions
(No. of University Teachers in 1990 as Treatment)

Dependent Variable:	Number of Patents								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Ring i	Ring 1	Ring 2	Ring 3	Ring 4	Ring 5	Ring 6	Ring 7	Ring 8	Ring 9
Treatment × 1995	-3.96	0.32	0.13	0.16	-0.05	-0.04	0.04	0.01	0.08
× Ring i	(-1.28)	(0.54)	(0.60)	(1.47)	(-0.61)	(-0.27)	(0.49)	(0.08)	(0.71)
Treatment × 1996	-4.67	-0.02	0.40*	0.17*	0.14	0.08	0.01	0.08	0.01
× Ring i	(-1.63)	(-0.05)	(1.95)	(1.69)	(0.96)	(0.78)	(0.17)	(0.94)	(0.09)
Treatment × 1997	-4.46*	-0.44	0.17	0.08	0.08	0.06	0.26	0.03	0.02
× Ring i	(-1.88)	(-1.04)	(1.04)	(1.08)	(0.89)	(0.64)	(1.64)	(0.53)	(0.43)
Treatment × 1998	-4.17*	-0.37	-0.08	-0.05	-0.18***	-0.08	0.03	-0.03	-0.07
× Ring i	(-1.81)	(-0.84)	(-0.61)	(-0.41)	(-2.79)	(-0.97)	(0.55)	(-0.55)	(-1.57)
Treatment × 2000	27.58*	2.39***	0.62***	0.16	0.10	0.22***	0.05	-0.08*	-0.01
× Ring i	(1.72)	(3.40)	(4.08)	(1.12)	(1.21)	(3.44)	(1.27)	(-1.74)	(-0.28)
Treatment × 2001	29.54***	2.92***	1.32***	0.92	0.05	0.00	0.12	0.07	0.13
× Ring i	(2.89)	(2.65)	(4.67)	(1.45)	(0.26)	(0.09)	(1.48)	(1.03)	(0.69)
Treatment × 2002	36.10***	11.09***	2.24***	0.77**	0.23	0.32***	0.34	-0.02	0.03
× Ring i	(3.40)	(3.45)	(2.77)	(2.00)	(1.22)	(5.01)	(1.06)	(-0.19)	(0.34)
Treatment × 2003	56.37***	14.37***	3.02***	1.70**	0.39	0.34*	0.53**	0.18	0.16
× Ring i	(4.86)	(2.66)	(2.86)	(2.25)	(0.94)	(1.85)	(2.25)	(0.97)	(1.11)
Treatment × 2004	75.26***	20.61**	4.28***	2.35**	0.74	0.24	0.55**	0.08	-0.07
× Ring i	(4.10)	(2.58)	(3.35)	(2.49)	(1.52)	(0.99)	(2.01)	(0.62)	(-0.45)
Treatment × 2005	104.91***	22.39***	5.15*	2.78*	1.59	1.39**	0.68	-0.26	-0.20
× Ring i	(3.67)	(3.22)	(1.94)	(1.78)	(1.20)	(2.30)	(1.00)	(-0.62)	(-0.50)
Treatment × 2006	117.54***	23.46***	11.62**	5.27**	2.17	1.48**	1.00	-0.05	-0.47
× Ring i	(3.86)	(4.23)	(2.14)	(2.24)	(1.46)	(2.10)	(1.31)	(-0.10)	(-0.92)
Treatment × 2007	146.65***	25.87***	15.10*	5.12*	1.11	1.17	0.05	-1.10*	-1.32**
× Ring i	(3.65)	(4.59)	(1.75)	(1.91)	(0.61)	(0.85)	(0.05)	(-1.77)	(-2.58)
Dependent Variable Mean	10.59	Observations	23920	Adj. R^2	0.548	Fixed Effects	Yes		

Notes: This table reports the estimates of the dynamic effects of university expansion on industry patents at different distances (rings). The estimates are used to plot Panel (a) of Figure 7. Year × Ring, Year × City, and City × Ring fixed effects are included in all regressions. The city-ring-specific pre-expansion time trend is removed for the dependent variable in all specifications. t statistics based on clustered standard errors at the city level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A7.B: The Dynamic Effects of University Expansion on Industry Innovation — Ring Regressions
(No. of University Students in 1990 as Treatment)

Dependent Variable:	Number of Patents								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Ring i	Ring 1	Ring 2	Ring 3	Ring 4	Ring 5	Ring 6	Ring 7	Ring 8	Ring 9
Treatment × 1995	-0.72	0.08	0.02	0.04*	-0.01	-0.01	0.01	-0.00	0.01
× Ring i	(-1.03)	(0.63)	(0.51)	(1.78)	(-0.68)	(-0.41)	(0.41)	(-0.14)	(0.61)
Treatment × 1996	-0.89	0.00	0.08*	0.03	0.02	0.01	0.00	0.01	0.00
× Ring i	(-1.35)	(0.04)	(1.71)	(1.52)	(0.78)	(0.65)	(0.02)	(0.64)	(0.01)
Treatment × 1997	-0.91*	-0.08	0.03	0.02	0.01	0.01	0.05	0.00	0.00
× Ring i	(-1.66)	(-0.89)	(0.87)	(1.00)	(0.69)	(0.49)	(1.43)	(0.21)	(0.37)
Treatment × 1998	-0.88*	-0.06	-0.02	-0.00	-0.04**	-0.02	0.00	-0.01	-0.01
× Ring i	(-1.72)	(-0.64)	(-0.74)	(-0.20)	(-2.12)	(-1.08)	(0.43)	(-0.66)	(-1.37)
Treatment × 2000	5.88*	0.47***	0.13***	0.04*	0.02	0.05***	0.01	-0.02	-0.00
× Ring i	(1.66)	(2.82)	(3.52)	(1.65)	(1.02)	(3.06)	(1.35)	(-1.40)	(-0.00)
Treatment × 2001	5.81**	0.62***	0.28***	0.21	0.02	0.00	0.03*	0.01	0.02
× Ring i	(2.45)	(2.65)	(4.41)	(1.60)	(0.42)	(0.14)	(1.73)	(0.74)	(0.54)
Treatment × 2002	7.27***	2.32***	0.47**	0.18**	0.06*	0.07***	0.08	-0.01	0.00
× Ring i	(2.90)	(3.09)	(2.59)	(2.34)	(1.69)	(4.39)	(1.17)	(-0.35)	(0.16)
Treatment × 2003	11.75***	3.15***	0.66***	0.39***	0.11	0.08**	0.12**	0.04	0.04
× Ring i	(3.97)	(2.77)	(2.90)	(2.73)	(1.37)	(2.23)	(2.56)	(1.08)	(1.31)
Treatment × 2004	15.30***	4.40**	0.93***	0.52***	0.17*	0.06	0.12**	0.02	-0.02
× Ring i	(3.38)	(2.56)	(3.52)	(2.71)	(1.75)	(1.09)	(2.22)	(0.53)	(-0.61)
Treatment × 2005	20.93***	4.69***	1.16**	0.63**	0.36	0.30**	0.15	-0.06	-0.04
× Ring i	(3.02)	(2.96)	(2.13)	(1.99)	(1.26)	(2.26)	(1.04)	(-0.70)	(-0.42)
Treatment × 2006	23.54***	4.88***	2.53**	1.18**	0.48	0.30*	0.21	-0.02	-0.11
× Ring i	(3.17)	(3.62)	(2.25)	(2.42)	(1.49)	(1.82)	(1.23)	(-0.19)	(-1.00)
Treatment × 2007	29.19***	5.34***	3.34*	1.19**	0.30	0.29	0.05	-0.20	-0.26**
× Ring i	(3.02)	(3.75)	(1.83)	(2.29)	(0.81)	(1.01)	(0.22)	(-1.39)	(-2.20)
Dependent Variable Mean	10.59	Observations	23920	Adj. R^2	0.526	Fixed Effects	Yes		

Notes: This table reports the estimates of the dynamic effects of university expansion on industry patents at different distances (rings). The estimates are used to plot Panel (b) of Figure 7. Year × Ring, Year × City, and City × Ring fixed effects are included in all regressions. The city-ring-specific pre-expansion time trend is removed for the dependent variable in all specifications. t statistics based on clustered standard errors at the city level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A8: Impact of University Expansion on Industry Innovation — Ring-level Regressions of Trend Break Model

	(1) Ring 1	(2) Ring 2	(3) Ring 3	(4) Ring 4	(5) Ring 5	(6) Ring 6	(7) Ring 7	(8) Ring 8	(9) Ring 9	(10) Ring 10	(11) Ring 1 - 10
Treatment × Trend	3.494^{***}										3.425^{***}
×After× 0.5km	(2.98)										(2.93)
Treatment × Trend		0.861^{***}									0.791^{***}
×After× 1km		(3.91)									(3.81)
Treatment × Trend			0.570^{**}								0.500^{*}
×After× 1.5km			(2.14)								(1.96)
Treatment × Trend				0.269^{***}							0.200^{***}
×After× 2km				(3.10)							(2.62)
Treatment × Trend					0.140^{**}						0.0710
×After× 2.5km					(2.02)						(1.18)
Treatment × Trend						0.120^{***}					0.0512
×After× 3km						(2.67)					(1.43)
Treatment × Trend							0.0862^{**}				0.0169
×After× 3.5km							(2.41)				(0.59)
Treatment × Trend								0.0468^{***}			-0.0225
×After× 4km								(2.70)			(-1.18)
Treatment × Trend									0.0334^{*}		-0.0358^{*}
×After× 4.5km									(1.88)		(-1.76)
Treatment × Trend										0.0693^{***}	
×After× 5km										(2.95)	
Observations	2392	2392	2392	2392	2392	2392	2392	2392	2392	2392	23920
City FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Year × Ring FE	No	No	No	No	No	No	No	No	No	No	Yes
Year × City FE	No	No	No	No	No	No	No	No	No	No	Yes
City × Ring FE	No	No	No	No	No	No	No	No	No	No	Yes
Dependent Variable Mean	114.50	26.42	13.04	8.26	5.71	4.37	3.34	2.49	1.82	2.22	18.21
Adj. R^2	0.584	0.660	0.565	0.632	0.457	0.456	0.455	0.383	0.433	0.212	0.607

Notes: This table reports the estimates of the slope change in the number of industry patents at different distances (rings) as a result of the university expansion, using the specification in Equation (3.8). The number of university students in 1990 is counted in 1,000, and it is used as the measure of treatment intensity. The trend-break model is used in all specifications. t statistics based on clustered standard errors at the city level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A9: Robustness Check — Ring-level Regressions up to 10 km

	Pre-expansion Time Trend Not Removed		Pre-expansion Time Trend Removed	
Dependent Variable	Number of Patents			
	(1)	(2)	(3)	(4)
Treatment × After × 0.5km	18.43***	18.41***	15.75***	15.76***
	(3.61)	(3.62)	(3.09)	(3.09)
Treatment × After × 1km	4.02***	4.00***	3.36***	3.36***
	(3.91)	(3.92)	(3.27)	(3.30)
Treatment × After × 1.5km	1.53***	1.51***	1.28**	1.28**
	(2.71)	(2.72)	(2.26)	(2.32)
Treatment × After × 2km	0.78***	0.76***	0.64**	0.64**
	(2.93)	(2.96)	(2.40)	(2.52)
Treatment × After × 2.5km	0.40**	0.38**	0.31*	0.31*
	(2.24)	(2.23)	(1.71)	(1.84)
Treatment × After × 3km	0.33***	0.30***	0.26**	0.26**
	(2.89)	(2.96)	(2.27)	(2.55)
Treatment × After × 3.5km	0.26**	0.23**	0.20*	0.20**
	(2.42)	(2.39)	(1.86)	(2.08)
Treatment × After × 4km	0.13***	0.11***	0.08*	0.09**
	(2.97)	(3.20)	(1.84)	(2.57)
Treatment × After × 4.5km	0.11**	0.08**	0.07	0.07*
	(2.15)	(2.17)	(1.34)	(1.88)
Treatment × After × 5km	0.15***	0.13***	0.11**	0.12***
	(3.00)	(2.94)	(2.25)	(2.74)
Treatment × After × 5.5km	0.10***	0.07***	0.06*	0.07***
	(3.15)	(3.26)	(1.95)	(2.93)
Treatment × After × 6km	0.06**	0.04*	0.03	0.03
	(2.12)	(1.84)	(0.95)	(1.55)
Treatment × After × 6.5km	0.07***	0.05***	0.04	0.04***
	(2.97)	(3.15)	(1.61)	(2.91)
Treatment × After × 7km	0.07**	0.04	0.03	0.04
	(2.40)	(1.64)	(1.22)	(1.49)
Treatment × After × 7.5km	0.05*	0.03	0.02	0.03
	(1.78)	(0.95)	(0.67)	(0.80)
Treatment × After × 8km	0.01	-0.02	-0.03**	-0.02**
	(0.53)	(-1.65)	(-2.56)	(-2.03)
Treatment × After × 8.5km	0.03**	0.00	-0.00	0.00
	(2.33)	(0.15)	(-0.44)	(0.03)
Treatment × After × 9km	0.12	0.10	0.09	0.09
	(1.37)	(1.08)	(1.00)	(1.04)
Treatment × After × 9.5km	0.02**	-0.00	-0.01	-0.00
	(2.22)	(-0.06)	(-0.76)	(-0.23)
Treatment × After × 10km	0.02*	-	-0.01	-
	(1.91)	-	(-0.44)	-
Observations	47840	47840	47840	47840
Treatment × Ring Dummies	Yes	No	Yes	No
Year × Ring FE	Yes	Yes	Yes	Yes
Year × City FE	No	Yes	No	Yes
City × Ring FE	Yes	Yes	Yes	Yes
Dependent Variable Mean	9.62	9.62	5.70	5.70
Adj. R^2	0.338	0.552	0.215	0.473

Notes: This table reports the estimates of the effects of university expansion on industry patents at different distances (rings) for up to 10 km. The city-ring-specific pre-expansion time trend is removed for the dependent variables in Columns (3) and (4). The number of university students in 1990 is counted in 1,000, and it is used as the measure of treatment intensity. t statistics based on clustered standard errors at the city level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A10.A: Spatial Decay of University Spillovers (Relative to the Effect on Ring 1)

	Pre-expansion Time Trend Not Removed		Pre-expansion Time Trend Removed	
Dependent Variable	Number of Patents			
	(1)	(2)	(3)	(4)
Treatment × After × 0.5km	-	-	-	-
	-	-	-	-
Treatment × After × 1km	-71.37***	-71.37***	-61.21***	-61.21***
	(-3.82)	(-3.82)	(-3.28)	(-3.28)
Treatment × After × 1.5km	-83.43***	-83.43***	-71.23***	-71.23***
	(-4.18)	(-4.18)	(-3.57)	(-3.57)
Treatment × After × 2km	-86.92***	-86.92***	-74.22***	-74.22***
	(-4.30)	(-4.30)	(-3.67)	(-3.67)
Treatment × After × 2.5km	-88.64***	-88.64***	-75.71***	-75.71***
	(-4.39)	(-4.39)	(-3.75)	(-3.75)
Treatment × After × 3km	-88.91***	-88.91***	-75.87***	-75.87***
	(-4.41)	(-4.41)	(-3.76)	(-3.76)
Treatment × After × 3.5km	-89.26***	-89.26***	-76.18***	-76.18***
	(-4.40)	(-4.40)	(-3.75)	(-3.75)
Treatment × After × 4km	-89.81***	-89.81***	-76.68***	-76.68***
	(-4.44)	(-4.44)	(-3.79)	(-3.79)
Treatment × After × 4.5km	-89.92***	-89.92***	-76.75***	-76.75***
	(-4.46)	(-4.46)	(-3.80)	(-3.80)
Treatment × After × 5km	-89.70***	-89.70***	-76.52***	-76.52***
	(-4.44)	(-4.44)	(-3.79)	(-3.79)
Treatment × After × 5.5km	-	-89.96***	-	-76.77***
	-	(-4.45)	-	(-3.80)
Treatment × After × 6km	-	-90.12***	-	-76.92***
	-	(-4.46)	-	(-3.80)
Treatment × After × 6.5km	-	-90.06***	-	-76.86***
	-	(-4.46)	-	(-3.81)
Treatment × After × 7km	-	-90.09***	-	-76.89***
	-	(-4.44)	-	(-3.79)
Treatment × After × 7.5km	-	-90.16***	-	-76.96***
	-	(-4.44)	-	(-3.79)
Treatment × After × 8km	-	-90.38***	-	-77.18***
	-	(-4.46)	-	(-3.81)
Treatment × After × 8.5km	-	-90.27***	-	-77.05***
	-	(-4.46)	-	(-3.81)
Treatment × After × 9km	-	-89.73***	-	-76.52***
	-	(-4.49)	-	(-3.83)
Treatment × After × 9.5km	-	-90.30***	-	-77.08***
	-	(-4.46)	-	(-3.80)
Observations	23920	47840	23920	47840
Year × Ring FE	Yes	Yes	Yes	Yes
Year × City FE	Yes	Yes	Yes	Yes
City × Ring FE	Yes	Yes	Yes	Yes
Dependent Variable Mean	18.21	9.62	10.91	5.86
Adj. R^2	0.570	0.563	0.482	0.475

Notes: This table reports the estimates of the effects of university expansion on industry patents at different distances (rings) for up to 5 km or 10 km. The reference group is ring 1. The estimates are used to plot Figure 10. The number of university teachers in 1990 is counted in 1,000. and it is used as the measure of treatment intensity. t statistics based on clustered standard errors at the city level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A10.B: Spatial Decay of University Spillovers (Relative to the Effect on Ring 1)

	Pre-expansion Time Trend Not Removed		Pre-expansion Time Trend Removed	
Dependent Variable	Number of Patents			
	(1)	(2)	(3)	(4)
Treatment ×	-	-	-	-
After × 0.5km	-	-	-	-
Treatment ×	-14.41***	-14.41***	-12.39***	-12.39***
After × 1km	(-3.19)	(-3.19)	(-2.74)	(-2.74)
Treatment ×	-16.90***	-16.90***	-14.47***	-14.47***
After × 1.5km	(-3.43)	(-3.43)	(-2.94)	(-2.93)
Treatment ×	-17.65***	-17.65***	-15.11***	-15.11***
After × 2km	(-3.50)	(-3.50)	(-3.00)	(-3.00)
Treatment ×	-18.03***	-18.03***	-15.45***	-15.45***
After × 2.5km	(-3.56)	(-3.56)	(-3.05)	(-3.05)
Treatment ×	-18.10***	-18.10***	-15.49***	-15.49***
After × 3km	(-3.58)	(-3.58)	(-3.06)	(-3.06)
Treatment ×	-18.17***	-18.17***	-15.55***	-15.55***
After × 3.5km	(-3.57)	(-3.57)	(-3.06)	(-3.06)
Treatment ×	-18.30***	-18.30***	-15.67***	-15.67***
After × 4km	(-3.60)	(-3.60)	(-3.08)	(-3.08)
Treatment ×	-18.33***	-18.33***	-15.68***	-15.68***
After × 4.5km	(-3.61)	(-3.61)	(-3.09)	(-3.09)
Treatment ×	-18.28***	-18.28***	-15.64***	-15.64***
After × 5km	(-3.60)	(-3.60)	(-3.08)	(-3.08)
Treatment ×	-	-18.34***	-	-15.69***
After × 5.5km	-	(-3.61)	-	(-3.09)
Treatment ×	-	-18.37***	-	-15.72***
After × 6km	-	(-3.61)	-	(-3.09)
Treatment ×	-	-18.36***	-	-15.71***
After × 6.5km	-	(-3.62)	-	(-3.09)
Treatment ×	-	-18.36***	-	-15.72***
After × 7km	-	(-3.60)	-	(-3.08)
Treatment ×	-	-18.38***	-	-15.73***
After × 7.5km	-	(-3.60)	-	(-3.08)
Treatment ×	-	-18.43***	-	-15.78***
After × 8km	-	(-3.62)	-	(-3.10)
Treatment ×	-	-18.41***	-	-15.76***
After × 8.5km	-	(-3.62)	-	(-3.10)
Treatment ×	-	-18.31***	-	-15.66***
After × 9km	-	(-3.64)	-	(-3.11)
Treatment ×	-	-18.41***	-	-15.76***
After × 9.5km	-	(-3.61)	-	(-3.09)
Observations	23920	47840	23920	47840
Year × Ring FE	Yes	Yes	Yes	Yes
Year × City FE	Yes	Yes	Yes	Yes
City × Ring FE	Yes	Yes	Yes	Yes
Dependent Variable Mean	18.21	9.62	10.59	5.70
Adj. R^2	0.560	0.552	0.479	0.473

Notes: This table reports the estimates of the effects of university expansion on industry patents at different distances (rings) for up to 5 km or 10 km. The reference group is ring 1. The estimates are used to plot Figure 10. The number of university students in 1990 is counted in 1,000. and it is used as the measure of treatment intensity. t statistics based on clustered standard errors at the city level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A11: Heterogeneity Analysis — Eastern, Central, and Western Regions

	Pre-expansion Time Trend Not Removed			Pre-expansion Time Trend Removed		
Dependent Variable	Number of Patents					
	(1)	(2)	(3)	(4)	(5)	(6)
	Eastern	Central	Western	Eastern	Central	Western
Treatment ×	25.44***	6.38***	5.01***	21.91***	5.14***	4.03***
After × 0.5km	(5.57)	(7.13)	(6.01)	(4.80)	(5.74)	(4.83)
Treatment ×	5.42***	1.20***	1.20***	4.66***	0.77**	0.83***
After × 1km	(4.78)	(3.43)	(5.12)	(4.12)	(2.20)	(3.54)
Treatment ×	1.91**	0.45***	0.46***	1.70**	0.21***	0.26**
After × 1.5km	(2.51)	(7.68)	(3.77)	(2.23)	(3.68)	(2.10)
Treatment ×	0.76**	0.39***	0.40*	0.65*	0.28***	0.33
After × 2km	(2.11)	(8.60)	(2.01)	(1.80)	(6.29)	(1.67)
Treatment ×	0.31	0.17***	0.09***	0.26	0.09***	0.04*
After × 2.5km	(1.30)	(7.89)	(4.33)	(1.08)	(4.33)	(2.03)
Treatment ×	0.22*	0.07***	0.13***	0.19	0.03**	0.10***
After × 3km	(1.69)	(4.63)	(6.07)	(1.45)	(2.19)	(4.77)
Treatment ×	0.11	0.08***	0.10*	0.09	0.06***	0.07
After × 3.5km	(0.84)	(5.13)	(1.84)	(0.69)	(3.69)	(1.35)
Treatment ×	-0.05	0.03**	0.05	-0.06	0.03**	0.02
After × 4km	(-0.88)	(2.42)	(1.50)	(-1.09)	(2.42)	(0.53)
Treatment ×	-0.07	0.00	0.02	-0.08	0.00	0.01
After × 4.5km	(-1.17)	(0.21)	(1.00)	(-1.19)	(0.15)	(0.38)
Treatment ×	-	-	-	-	-	-
After × 5km	-	-	-	-	-	-
Observations	10920	8320	4550	10920	8320	4550
Year × Ring FE	Yes	Yes	Yes	Yes	Yes	Yes
Year × City FE	Yes	Yes	Yes	Yes	Yes	Yes
City × Ring FE	Yes	Yes	Yes	Yes	Yes	Yes
Dependent Variable Mean	30.55	6.41	10.69	19.53	2.97	6.09
Adj. R^2	0.582	0.697	0.675	0.491	0.533	0.557

Notes: This table reports the estimated effects of university expansion on industry patents across different regions in China. The Eastern, Central and Western regions are divided according to the 7th "Five-Year Plan for the National Economic and Social Development" of China. The city-ring-specific pre-expansion time trend is removed for the dependent variables in Columns (4)-(6). The number of university students in 1990 is counted in 1,000, and it is used as the measure of treatment intensity. t statistics based on clustered standard errors at the city level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A12: Heterogeneity Analysis — Industries with High, Medium, and Low Human Capital Intensity

	Pre-expansion Time Trend Not Removed			Pre-expansion Time Trend Removed		
Dependent Variable	Number of Patents					
	(1)	(2)	(3)	(4)	(5)	(6)
	High	Medium	Low	High	Medium	Low
Treatment ×	13.78***	3.03***	0.81***	11.72***	2.44***	0.62***
After × 0.5km	(3.11)	(4.39)	(4.36)	(2.64)	(3.53)	(3.33)
Treatment ×	2.95***	0.71***	0.22***	2.48***	0.55**	0.16***
After × 1km	(3.70)	(3.13)	(3.67)	(3.11)	(2.40)	(2.78)
Treatment ×	1.01**	0.25***	0.11**	0.86*	0.17*	0.09
After × 1.5km	(2.30)	(2.83)	(2.05)	(1.94)	(1.94)	(1.65)
Treatment ×	0.36**	0.18*	0.06***	0.28**	0.14	0.05***
After × 2km	(2.59)	(1.87)	(3.82)	(2.02)	(1.50)	(3.31)
Treatment ×	0.10	0.05	0.03**	0.06	0.03	0.02
After × 2.5km	(1.13)	(1.24)	(2.02)	(0.65)	(0.67)	(1.51)
Treatment ×	0.10	0.04	0.02*	0.07	0.03	0.02
After × 3km	(1.24)	(1.58)	(1.82)	(0.89)	(1.20)	(1.49)
Treatment ×	0.03	0.02	0.02*	0.01	0.01	0.02*
After × 3.5km	(0.35)	(0.94)	(1.92)	(0.14)	(0.58)	(1.76)
Treatment ×	-0.07	0.02	0.01*	-0.08	0.01	0.00
After × 4km	(-1.31)	(0.56)	(1.69)	(-1.55)	(0.43)	(1.25)
Treatment ×	-0.07	-0.01	0.01	-0.08	-0.01	0.01
After × 4.5km	(-1.40)	(-0.55)	(1.18)	(-1.45)	(-0.61)	(1.21)
Treatment ×	-	-	-	-	-	-
After × 5km	-	-	-	-	-	-
Observations	23660	23660	23660	23660	23660	23660
Year × Ring FE	Yes	Yes	Yes	Yes	Yes	Yes
Year × City FE	Yes	Yes	Yes	Yes	Yes	Yes
City × Ring FE	Yes	Yes	Yes	Yes	Yes	Yes
Dependent Variable Mean	14.13	3.71	1.19	8.37	2.00	0.65
Adj. R^2	0.462	0.721	0.689	0.401	0.612	0.555

Notes: This table reports the estimated effects of university expansion on industry patents across industries with different human capital intensity. We define high human capital intensity industry as the industries that rank among the top one-third in the college employee ratio, medium as the middle one-third, and low as the rest. The industry college employee ratio is calculated as the percentage of workers with a college education and above using the 2004 ASIF. The city-ring-specific pre-expansion time trend is removed for the dependent variables in Columns (4)-(6). The number of university students in 1990 is counted in 1,000, and it is used as the measure of treatment intensity. t statistics based on clustered standard errors at the city level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A13: Summary Statistics of Firm Characteristics

Year	(1) No. of Firms	(2) New Product	(3) Output	(4) Fixed Assets	(5) SOE	(6) Firm Age	(7) Employment
1998	125954	3934.99	33653.85	45043.08	0.29	14.55	348.68
1999	126528	4515.49	36260.19	49413.91	0.26	14.71	362.18
2000	125204	5411.33	37641.48	57134.02	0.22	14.34	336.75
2001	134677	6311.51	37539.98	60340.50	0.18	12.73	313.96
2002	146290	7186.15	37347.27	65840.98	0.15	12.06	301.97
2003	159423	8216.19	33295.39	76591.23	0.11	10.91	287.69
2004	231811	–	31756.00	72843.79	0.08	8.64	232.73
2005	224051	9814.97	36728.77	91410.37	0.06	8.78	249.65
2006	247992	11814.78	37504.38	101072.40	0.05	8.64	237.28
2007	276058	12518.99	37743.88	110539.40	0.03	8.38	229.20

Notes: Column (1) reports the number of firms in each year. Columns (2)-(7) report the means of new product value, output, fixed assets, SOE status, firm age and employment at the firm level. New product, output, and fixed assets are counted in 1,000 yuan in the year 1998 value. Information on new product value in 2004 is not available.

Table A14: Effects of University Expansion on New Product Ratio — Ring-level Regressions

Dependent Variable	New Product Ratio			
	(1)	(2)	(3)	(4)
After Dummy	2000	2002	2004	2006
Treatment ×	8.47e-04	9.85e-04	1.30e-03	1.36e-03
After × 0.5km	(1.58)	(1.57)	(1.60)	(1.58)
Treatment ×	5.77e-04*	6.99e-04*	9.03e-04**	9.59e-04*
After × 1km	(1.95)	(1.96)	(2.03)	(1.87)
Treatment ×	3.96e-04***	4.67e-04***	6.16e-04***	6.44e-04***
After × 1.5km	(3.14)	(3.21)	(3.13)	(2.97)
Treatment ×	2.91e-04***	3.35e-04***	4.12e-04***	4.10e-04***
After × 2km	(2.89)	(2.87)	(3.27)	(2.89)
Treatment ×	3.60e-04**	4.41e-04**	4.13e-04***	3.92e-04***
After × 2.5km	(2.44)	(2.56)	(3.33)	(3.16)
Treatment ×	2.54e-04***	3.10e-04***	3.97e-04***	4.06e-04***
After × 3km	(3.18)	(3.41)	(3.99)	(3.79)
Treatment ×	1.56e-04**	1.88e-04**	2.63e-04***	3.02e-04***
After × 3.5km	(2.34)	(2.47)	(2.90)	(2.91)
Treatment ×	1.60e-04*	1.92e-04*	2.99e-04**	3.62e-04**
After × 4km	(1.74)	(1.89)	(2.35)	(2.42)
Treatment ×	1.77e-04**	2.01e-04**	2.69e-04***	2.96e-04***
After × 4.5km	(2.50)	(2.48)	(2.67)	(2.97)
Treatment ×	1.45e-04*	1.77e-04*	2.74e-04**	3.35e-04**
After × 5km	(1.95)	(1.94)	(2.16)	(2.57)
Observations	1196263	996185	759980	589233
Year × Ring FE	Yes	Yes	Yes	Yes
Year × City FE	Yes	Yes	Yes	Yes
City × Ring FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Control Variables	Yes	Yes	Yes	Yes
Dependent Variable Mean	0.034	0.035	0.035	0.037
Adj. R^2	0.091	0.097	0.107	0.105

Notes: This table reports the estimated effects of university expansion on firms' new product ratio using the number of university students in 1990 as the proxy for treatment intensity. Dependent variable is firm-level new product ratio. Columns (1)–(4) report the triple-differences estimates. The after dummy equals 1 if year is 2000 or after, 2002 or after, 2004 or after, or 2006 or after in Columns (1), (2), (3), and (4), respectively. The after dummy equals 0 for years before 2000 for all four columns. Observations in the years in which the after dummy is not defined are dropped. The reference group is the firms outside 10 km of universities. Control variables include firm age, fixed assets, SOE status, and employment size. Data on new product in 2004 is not available. The number of university students in 1990 is counted in 1,000. t statistics based on clustered standard errors at the city level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A15: Effects of University Expansion on New Product Ratio — Ring-level Regressions with Firm Fixed Effects

Dependent Variable	New Product Ratio			
	(1)	(2)	(3)	(4)
After Dummy	2000	2002	2004	2006
Treatment ×	8.08e-04**	9.10e-04*	1.82e-03***	2.20e-03**
After × 0.5km	(1.98)	(1.81)	(3.14)	(2.32)
Treatment ×	4.45e-04**	7.42e-04***	1.43e-03***	1.79e-03***
After × 1km	(2.38)	(3.25)	(6.46)	(2.66)
Treatment ×	6.44e-04***	9.24e-04***	1.12e-03***	1.15e-03**
After × 1.5km	(3.60)	(3.63)	(2.97)	(2.36)
Treatment ×	6.24e-04***	5.92e-04**	8.38e-04***	7.43e-04**
After × 2km	(3.01)	(2.17)	(3.28)	(2.46)
Treatment ×	5.63e-04***	7.71e-04**	1.41e-03***	1.04e-03**
After × 2.5km	(2.94)	(2.39)	(3.10)	(2.11)
Treatment ×	2.11e-04	2.26e-04	4.53e-04	3.28e-04
After × 3km	(1.13)	(1.22)	(1.40)	(0.77)
Treatment ×	3.45e-04*	4.05e-04*	3.12e-04	5.03e-04
After × 3.5km	(1.93)	(1.90)	(0.68)	(0.96)
Treatment ×	1.12e-04	3.08e-04	3.79e-04	-6.87e-05
After × 4km	(-0.57)	(-1.24)	(-1.07)	(-0.15)
Treatment ×	2.45e-04	5.87e-04	1.07e-03	1.05e-03
After × 4.5km	(-0.52)	(-0.90)	(-1.26)	(-1.18)
Treatment ×	2.99e-04*	2.53e-04	1.50e-04	3.57e-04
After × 5km	(1.71)	(0.87)	(0.34)	(0.54)
Observations	1099149	895751	668829	498355
Year × Ring FE	Yes	Yes	Yes	Yes
Year × City FE	Yes	Yes	Yes	Yes
City × Ring FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
Control Variables	Yes	Yes	Yes	Yes
Dependent Variable Mean	0.035	0.036	0.038	0.037
Adj. R^2	0.541	0.541	0.572	0.589

Notes: This table reports the estimated effects of university expansion on firms' new product ratio using the number of university teachers in 1990 as the proxy for treatment intensity, with firm fixed effects. The dependent variable is firm-level new product ratio. Columns (1)–(4) report the triple-differences estimates. The after dummy equals 1 if year is 2000 or after, 2002 or after, 2004 or after, or 2006 or after in Columns (1), (2), (3), and (4), respectively. The after dummy equals 0 if year is before 2000 for all four columns. Observations in the years in which the after dummy is not defined are dropped. The reference group is the firms outside 10 km of universities. Control variables include firm age, fixed assets, SOE status, and employment size. The number of university teachers in 1990 is counted in 1,000 t statistics based on clustered standard errors at the city level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A16: Heterogeneity Analysis - Industries with High, Medium, and Low Human Capital Intensity and SOE versus Non-SOE

Dependent Variable	New Product Ratio				
	(1) High	(2) Medium	(3) Low	(4) SOE	(5) Non-SOE
Treatment ×	9.49e-04	5.79e-04*	2.62e-04	1.49e-04	8.61e-04
After × 0.5km	(1.59)	(1.75)	(0.87)	(0.66)	(1.52)
Treatment ×	8.13e-04*	2.93e-04	2.32e-04*	2.03e-04	6.31e-04**
After × 1km	(1.92)	(1.58)	(1.95)	(1.37)	(2.01)
Treatment ×	6.16e-04***	2.93e-04***	3.76e-05	1.65e-04**	4.25e-04***
After × 1.5km	(2.91)	(3.37)	(0.73)	(2.22)	(3.04)
Treatment ×	3.30e-04*	1.93e-04**	1.66e-04***	1.12e-04	3.32e-04***
After × 2km	(1.74)	(2.39)	(2.70)	(1.24)	(3.22)
Treatment ×	6.04e-04*	1.41e-04**	1.61e-04***	3.14e-04***	3.75e-04**
After × 2.5km	(1.68)	(2.03)	(3.53)	(3.60)	(2.34)
Treatment ×	3.53e-04**	2.10e-04**	5.70e-05	2.25e-04**	2.86e-04***
After × 3km	(2.43)	(2.29)	(1.03)	(2.32)	(3.35)
Treatment ×	1.73e-04	1.09e-04	7.19e-05	2.55e-05	2.09e-04**
After × 3.5km	(1.23)	(1.16)	(1.14)	(0.23)	(2.49)
Treatment ×	1.27e-04	3.26e-05	1.62e-04***	-4.30e-05	2.19e-04*
After × 4km	(0.47)	(0.47)	(3.07)	(-0.35)	(1.88)
Treatment ×	2.35e-04**	-1.66e-05	2.13e-04***	-1.40e-04	2.43e-04***
After × 4.5km	(2.52)	(-0.13)	(3.86)	(-1.54)	(4.61)
Treatment ×	-4.94e-05	1.50e-04	1.73e-04**	5.77e-05	1.98e-04**
Observations	394427	385023	456632	136171	1060023
Year × Ring FE	Yes	Yes	Yes	Yes	Yes
Year × City FE	Yes	Yes	Yes	Yes	Yes
City × Ring FE	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes
Control Variables	Yes	Yes	Yes	Yes	Yes
Dependent Variable Mean	0.059	0.025	0.020	0.046	0.033
Adj. R^2	0.119	0.057	0.049	0.111	0.096

Notes: Columns (1)–(3) report the estimated effects of university expansion on firms' new product ratio across industries with different human capital intensity. We define high human capital intensity industry as the industries that rank among the top one-third in the college employee ratio, medium as the middle one-third, and low as the rest. The industry college employee ratio is calculated as the percentage of workers with a college education and above using the 2004 ASIF. Columns (4) and (5) report the estimates of the effects of university expansion on firms' new product ratio for SOEs and non-SOEs separately. The number of university students in 1990 is counted in 1,000, and it is used as the treatment intensity. Control variables include firm age, fixed assets, SOE status, and employment size. The reference group consists of the firms outside 10 km of universities. t statistics based on clustered standard errors at the city level are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix B: Theoretical Model

The impact of universities on local innovation can be mediated through a collection of channels, such as an increased supply of human capital, knowledge spillovers, or a direct demand effect (Valero and Van Reenen 2019). Most channels, such as the human capital channel and demand effect, operate at a broad geographic scale. For example, the human capital channel usually operates at the city level because workers are mobile within a city. The geographic scope of knowledge spillovers, however, is often limited. We highlight below the specific channels through which the impact of universities is especially pronounced at close geographic distances (within 2-3 km in general).

First, areas with better access to universities benefit from improved chances of collaborating with universities to convert university-based knowledge to commercial products. The development of the Founder Group in Zhongguancun (ZGC) Science Park is a typical example. The Founder Group, established by Peking University in 1986, is now a major Chinese technology conglomerate. The company's take-off benefited tremendously from Professor Xuan Wang at Peking University, who is known as the "Father of Chinese Character Laser Typesetting." His laser typesetting system allowed the Founder Group to earn its first pot of gold in the early 1990s. Between 2000 and 2007, the Founder Group published 232 invention patents, with 86 percent in collaboration with Peking University.

Second, universities may disproportionately benefit firms in close proximity through knowledge transfers. Areas close to universities enjoy convenient access to fundamental background knowledge and frontier technologies produced by university-based experts and professionals. Those factors are key drivers of innovation. Theory-based fundamental knowledge is the essential cornerstone of applicable innovations. Frontier technologies—grown out from the development of the fundamentals—lead to new commercializable product varieties as in Romer (1990) or upgrading of existing products through Schumpeterian creative destruction (Aghion and Howitt 1992; Grossman and Helpman 1991). The tacit part of knowledge and technologies requires lengthy face-to-face communications to disseminate. Locating close to universities allows nearby inventors to attend university workshops, seminars, and conferences and offers abundant opportunities for face-to-face interactions.

Third, proximity to universities allows firms to establish and foster strong professional and informational networks. Technology advancement is fast-evolving and subject to uncertain dynamics. Maintaining formal and informal operational links with universities and other research institutions to receive ceaseless updates on new information allows innovative firms to be at the front of technology development. A salient example is Baidu, Inc. Upon returning from Silicon Valley, the founder and CEO of Baidu, Yanhong Li, chose ZGC Science Park to develop his Chinese search engine empire. As he revealed in an interview, the location advantage of ZGC allows the company to maintain strong ties with experts at nearby universities, including Peking University from which he graduated (Zhao 2018). This idea is closely related to the networking benefits in the advertising industry as emphasized in Arzaghi and Henderson (2008).

In sum, firms in close proximity to universities benefit from improved collaboration opportunities, knowledge transfers, and information networks. However, we note a fundamental distinction between direct collaborations and the latter two channels. Collaboration benefits do not constitute spillovers because universities would internalize the benefits. We refer to the latter two channels as knowledge spillovers, which is the main focus of this paper. We also explore the collaboration channel quantitatively by treating innovative firms in direct collaboration with universities differently in our empirical analysis.

Next, we outline a simple conceptual framework to formalize the identification of knowledge spillovers by drawing on the localized nature of knowledge spillovers documented in the literature. Note again that we focus on variations within close geographic distances (within 2-3 km in general) to identify the fast spatial decay of knowledge spillovers. In this framework, the number of new ideas, NI, is assumed to be a function of the existing knowledge stock, A, and the number of researchers, R, who spend time producing them:

$$NI = f(A, R). \tag{1.11}$$

Conceptually, we specify the production function for each firm, i, but i is suppressed for simplicity. The number of new products, NP, is assumed to be a function of new ideas, NI, and the necessary facility, equipment, and personnel, X, to convert the new ideas into new products:

$$NP = g(NI, X). \tag{1.12}$$

We further assume that the knowledge stock, A, is affected by a nearby university's scale or its innovation capacity, U, through the channel of knowledge spillovers as well as the distance to the university, D, which captures the sharp spatial attenuation of knowledge spillovers:

$$A = a(U, D). \tag{1.13}$$

If nearby universities experience an increase in innovation capacity and generate more knowledge for sharing, the knowledge stock for nearby firms will increase. If a firm is closer to universities spatially, the firm has better access to university knowledge and receives a larger impact when the universities experience a knowledge boom. Therefore, we have $\frac{\partial a}{\partial U} > 0$, $\frac{\partial a}{\partial D} < 0$, and $\frac{\partial^2 a}{\partial U \partial D} < 0$.

Knowledge spillover is not the only channel through which universities affect local innovation. For instance, the number of researchers, R, could also be a function of local universities' scale, U:

$$R = r(U). \tag{1.14}$$

On a broad geographic scale, the number of researchers could also be a function of the geographic distance to the university. For instance, better university access may increase the probability that local young people attend a university, become researchers, and seek work in the same city (Card 1995). However, in this paper, we restrict our attention to narrow geographic scopes of 2-3 km, within which the number of available researchers to firms are unlikely to be subject to spatial attenuation.[48] Hence, we assume away the role of distance in driving the number of

[48] The argument is consistent with the general consensus in the literature that it is easier to "move" labor than to "move" ideas (Rosenthal and Strange 2001; Ganguli et al. 2020). The chances of meetings and conversations that enable idea exchanges are significantly reduced even at modest distances (Arzaghi and Henderson 2008). Yet, labor market benefits are realized at a large geographic scope—usually within the same commuting zones (Combes and Gobillon 2015).

available researchers in nearby firms.[49]

Based on the conceptual setup, it is easy to see that an increase in local university scale impacts the creation of new ideas and new products through either knowledge spillovers or the labor market channel. However, a further difference of the university impact along the spatial dimension helps tease out the labor market mechanism and highlight the role of knowledge spillovers. To see this clearly, we assume linearity for all functional forms and write the determinants of new ideas at "close" and "far" distances as follows:

$$NI_{D(close)} = A_{D(close)} + R_{D(close)} = \alpha_{D(close)}U + \beta U, \tag{1.15}$$

and

$$NI_{D(far)} = A_{D(far)} + R_{D(far)} = \alpha_{D(far)}U + \beta U, \tag{1.16}$$

where $NI_{D(close)}$ stands for the creations of new ideas at firms located sufficiently close to a university and $NI_{D(far)}$ stands for new ideas at firms located relatively far from the university. They are determined by the knowledge stock and the number of researchers at respective locations, indexed by $A_{D(close)}$, $A_{D(far)}$, $R_{D(close)}$, and $R_{D(far)}$. $\alpha_{D(close)}$, $\alpha_{D(far)}$, and β are the corresponding parameters that link $A_{D(close)}$, $A_{D(far)}$, $R_{D(close)}$, and $R_{D(far)}$ to U.

Since the number of available researchers to firms are not subject to spatial attenuation, as discussed earlier, we have $R_{D(close)} = R_{D(far)} = \beta U$. A comparison of the impact of universities for locations that are close and far from universities gives us the following.

$$NI_{D(close)} - NI_{D(far)} = A_{D(close)} - A_{D(far)} = \left[\alpha_{D(close)} - \alpha_{D(far)}\right]U. \tag{1.17}$$

Therefore, any differences in the impact of universities across various close-range spatial distances can be attributed to the difference in A—the varying degrees of knowledge spillovers in promoting nearby firms' innovation activities. We adopt a triple-differences model to highlight this variation in our empirical analysis.

[49]Essentially, the labor market channel should not operate in such localized geographic scales. However, the assumption does not preclude that, in equilibrium, firms closer to universities benefit more from knowledge spillovers and disproportionately hire more university graduates as researchers. This hiring is, in fact, likely if knowledge stock and the number of researchers are complementary.

Chapter 2

Analysis of Large Real Estate Prices Data: A High-Order Spatiotemporal Autoregression Approach

2.1 Introduction

Consider the real estate prices, collected unevenly over space and time, and arrived sequentially in time order. Suppose there are overall N spatial units involved in the data, and the observations on their response variable Y and explanatory variables X are denoted as $(Y_i, X_i), i = 1, \ldots, N$. Note that i indexes both the spatial unit and the time order. Let t_i be the time at which the ith spatial unit is observed. Note that in this type of data, it is typical that none of the spatial units are observed more than once during the time of study, e.g., none of the residential units are sold more than once within one year. Such a data is referred to in this paper as the *spatiotemporal data*, to distinguish from the regular panel or longitudinal data where a set of spatial units are repeatedly observed a number of times.[1]

Using spatiotemporal data for analyzing real estate prices has two main advan-

[1]In the real estate price literature, the term "spatiotemporal data" has been used in the literature to mean any data that are collected over space and time, which include the type of data just described (Pace et al. 1998a, 2000; Sun et al. 2005), the standard panel or longitudinal data (Holly et al. 2010, 2011), and the unbalanced panel data.

tages. First, compared with analyzing the real estate prices data by aggregating individual transactions to form a panel, it does not suffer from the loss of information as each individual transaction is incorporated. Also, aggregating the individual data leads to the ecological inference problem and *modifiable areal unit problem*,[2] as mentioned in Anselin and Cho (2002). Using spatiotemporal data instead of aggregated data alleviates these problems. Second, compared with traditional hedonic housing prices models, which usually do not consider the spatial and temporal information,[3] the usage of the additional spatial and temporal information allows researchers to capture interaction effects between transactions along both space and time dimensions. The richness of information in the spatiotemporal data also gives researchers a chance to control for cluster and time heterogeneity in the model.

Despite of these merits, spatial econometric theories and methods seem lag behind for sophisticated analyses of spatiotemporal data. Among a few papers which do use spatiotemporal data to model housing prices, Pace et al. (1998a) and Pace et al. (2000) propose a model with two different weight matrices which capture the interaction effects in space and time dimensions separately. They show that their model fits the housing prices data significantly better than the traditional hedonic model. Sun et al. (2005) add another layer of spatial effects to the model, which they call “building effects". Nappi-Choulet Pr and Maury (2009) use a similar model to analyze the housing market in Paris. Additionally, they propose a hybrid method for incorporating a temporal regime switch into the model, highlighting the fact that there may exist temporal heterogeneity. Baltagi et al. (2015) adopt a model with nested random effects to capture the neighborhood spillover effects in Paris. Anselin and Lozano-Gracia (2008) propose a hedonic house price model with spatial effects to evaluate the effect of improved air quality. Dorsey et al. (2010) adopt the hedonic house price model with spatial effects to construct better housing price indexes. However, the literature on spatiotemporal data lacks discus-

[2]Ecological inference problem describes the situation when the conclusion drawn from the data at an aggregated fails to explain individual level behaviour. The modifiable areal unit problem is about the proper spatial scale of analysis. Both the magnitude and sign of the spatial interaction effects can change when researchers aggregate the data at different levels.

[3]There is a large literature on hedonic house price studies. As mentioned in Maclennan (2012), “For almost four decades hedonic house price studies have been used to identify the economic significance of different, distinctive attributes of housing."

sions on the ways to incorporate 'cluster' and 'time segment' effects into the model and the ways of tackling the issues arising from the estimation of these effects when they are treated as fixed effects. Also, researchers usually assume homoskedasticity in the error term, which can potentially lead to inconsistency of the estimates, as housing price observations are typically heteroskedastic.

In this paper, we propose a high order **s**patio**t**emporal **a**uto**r**egressive (STAR) model for analyzing the spatiotemporal real estate prices data. By utilizing the spatial and temporal information in the data, our model is able to capture the spatiotemporal interaction effects between housing transactions. We develop methods to consistently estimate the spatiotemporal effects, at the same time allowing for various kinds of unobserved clusters and/or time segments fixed effects to control for cluster and time heterogeneity. We first follow the well-known *quasi maximum likelihood* (QML) method to estimate the model when errors are homoskedastic and the number of clusters (C) and time segments (T) are fixed. When C and/or T grow with N, an *adjusted quasi score* (AQS) method is proposed to deal with the *incidental parameters problem* (Neyman and Scott, 1948). Further more, potential *heteroskedasticity* in idiosyncratic errors can invalidate the QML and AQS estimates. To deal with this issue, we propose a *robust adjusted quasi score* (RAQS) method. Consistency and asymptotic normality of the QML, AQS and RAQS estimators are established, their validity are critically discussed, and instructions are clearly given on their practically implementations.

We present an empirical application of the proposed methods utilizing the housing transaction data in Beijing, to study the spatiotemporal interaction effects. We find strong evidence of the existence of the positive interaction effects. These effects take place through the interaction within communities and neighbours outside of communities.[4] However, we find that the estimation of the interaction effects are sensitive to controlling for cluster heterogeneity at the community level. It's very likely for us to severely underestimate the interaction effects within communities, and overestimate the interaction effects from neighbours that are outside of com-

[4]In this paper, the term "community" represents *xiao qu* in Chinese, which is usually a relatively independent residential area with walls.

munities if we do not allow for community heterogeneity. When we further allow for community heterogeneity together with heteroskedasticity in the error terms, the similar patterns still remain, but the extent to which we over/underestimate the interaction effects becomes smaller.

Our study contributes to the literature in threefold. First, controlling for various kinds of fixed effects have become a standard in empirical research, as the unobserved heterogeneity can be a source of endogeneity. To the best of our knowledge, this is the first paper that develops a series of methods to include different kinds of unobserved cluster and time segment fixed effects (especially when C and/or T are large) in the STAR model, and at the same time give consistent estimates of spatiotemporal interaction effects in both spatial lag and spatial error terms. Second, researchers have documented the existence of heteroskedasticity in hedonic models. The heteroskedasticity can be caused by dwelling age (Stevenson 2004; Goodman and Thibodeau 1995, 1997). It is well known that in simple linear models, OLS estimators are still consistent under heteroskedasticity, but it is not necessary the case in spatial econometrics models. The existence of heteroskedasticity can potentially make the estimation of the spatiotemporal interaction effects inconsistent. The RAQS estimators we proposed are able to consistently estimate the spatial parameters against unknown heteroskedasticity. Third, estimating the interaction effects (or spillover effects or price reference effects) is a very important topic in the housing literature, as it is closely related to people's decision on house-purchase and mortgage defaults (Campbell et al. 2011; Gerardi et al. 2012; Honkanen and Schmidt 2017; Gupta 2019). By applying our model to real estate prices data in Beijing, we show that estimates of the interaction effects are affected by taking into account the community fixed effects and the heteroskedasticity in the idiosyncratic error terms. Additionally, our detailed discussion on the construction the space-time connectivity matrices also contributes the literature on modelling interaction effects in the housing market.

The rest of the paper goes as follows. Section 2 introduces explicitly the type of data that we deal with in this paper, and the high-order STAR model with additive cluster and time segment fixed effects for analyzing these data. Section 3 introduces the QML estimation and inference of the model with iid errors and fixed (small) C and T. Section 4 considers the AQS estimation and inference of the model with large C and/or T. Section 5 considers the AQS estimation and inference when there exists unknown heteroskedasticity in the errors. Section 6 gives an extensive discussion on the construction of space-time connectivity matrices in the context of real estate prices. Section 7 presents the Monte Carlo results. Section 8 presents an empirical application utilizing real estate prices data in Beijing and Section 9 concludes the paper. Some technical details and additional results are collected in the appendix.

2.2 The High-Order Spatiotemporal Autoregressive Model

For ease of exposition, we begin by assuming that the N observations fall into T different time segments separated by the occurrence of common shocks, and they are classified into C clusters due to certain housing grouping/clustering features. As housing transactions occur in time sequence, they are not grouped naturally into clusters. To facilitate the subsequent discussions, the basic data structure is illustrated in the table below.

To identify which cluster that an observation falls into, we define an $N \times C$ *cluster-membership* matrix M_C such that its (i,c)th element equals 1 if the ith observation belongs to cth cluster, and equals to 0 otherwise, for $i = 1,\dots,N$ and $c = 1,\dots,C$. Similarly, we define an $N \times T$ matrix M_T such that its (i,t)th element equals 1 if the ith observation falls into tth time segment, and equals to 0 otherwise, for $i = 1,\dots,N$ and $c = 1,\dots,C$. In the empirical application, we use the real estate prices data in Beijing in 2015. The time segments can be weeks, months, or quarters, and the clusters can be districts or communities. We discuss the detailed ways of specifying clusters and time segments in Section 2.8.2.

The Structure of Spatiotemporal Data.

Obs.	Y	Cluster	Transaction Time	Time Segment	X_1	$\cdots$	X_k
1	Y_1	3	12.01.2015	1	X_{11}	$\cdots$	X_{k1}
2	Y_2	1	13.01.2015	1	X_{12}	$\cdots$	X_{k2}
3	Y_3	6	15.01.2015	1	X_{13}	$\cdots$	X_{k3}
4	Y_4	10	16.01.2015	1	X_{14}	$\cdots$	X_{k4}
5	Y_5	36	17.01.2015	1	X_{15}	$\cdots$	X_{k5}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
491	Y_{491}	10	20.05.2015	2	$X_{1,491}$	$\cdots$	$X_{k,491}$
492	Y_{492}	5	23.05.2015	2	$X_{1,492}$	$\cdots$	$X_{k,492}$
493	Y_{493}	23	25.05.2015	2	$X_{1,493}$	$\cdots$	$X_{k,493}$
494	Y_{494}	33	26.05.2015	2	$X_{1,494}$	$\cdots$	$X_{k,494}$
495	Y_{495}	3	29.05.2015	2	$X_{1,495}$	$\cdots$	$X_{k,495}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
991	Y_{991}	15	20.11.2015	3	$X_{1,991}$	$\cdots$	$X_{k,991}$
992	Y_{992}	12	23.11.2015	3	$X_{1,992}$	$\cdots$	$X_{k,912}$
993	Y_{993}	16	25.11.2015	3	$X_{1,993}$	$\cdots$	$X_{k,993}$
994	Y_{994}	20	26.11.2015	3	$X_{1,994}$	$\cdots$	$X_{k,994}$
995	Y_{995}	6	29.11.2015	3	$X_{1,995}$	$\cdots$	$X_{k,995}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

We propose a high-order *spatiotemporal autoregressive* (STAR) model with cluster and time segment fixed effects:

$$A_N(\lambda)Y_N = X_N\beta + M_C\mu + M_T\alpha + V_N,\;\; B_N(\rho)V_N = \varepsilon_N, \tag{2.1}$$

where Y_N is an $N \times 1$ vector of response values, X_N an $N \times k$ matrix containing the values of exogenous regressors, V_N an $N \times 1$ vector of disturbances, and ε_N an $N \times 1$ vector of idiosyncratic errors. β is a $k \times 1$ vector of regression coefficients, μ a $C \times 1$ vector of cluster fixed effects, and α a $T \times 1$ time segment fixed effects. $A_N(\lambda)$ and $B_N(\rho)$, are defined as

$$A_N(\lambda) \equiv A_N(\mathbb{W}_\ell, \lambda) = I_N - \textstyle\sum_{j=1}^{p} \lambda_j W_{\ell j}, \tag{2.2}$$

$$B_N(\rho) \equiv B_n(\mathbb{W}_e, \rho) = I_N - \textstyle\sum_{k=1}^{q} \rho_j W_{ek}, \tag{2.3}$$

which are $N \times N$ matrices capturing, respectively, the *space-time lag* (STL) dependence of order p, and the *space-time error* (STE) dependence of order q. The vectors, $\lambda = \{\lambda_1, \ldots, \lambda_p\}'$ and $\rho = \{\rho_1, \ldots, \rho_q\}'$, are the spatiotemporal parameters. The $N \times N$ matrices $\{W_{\ell j}\}$ and $\{W_{ek}\}$, are the given space-time weight matrices capturing various space-time effects, which are denoted collectively as $\mathbb{W}_\ell = \{W_{\ell 1}, \cdots, W_{\ell p}\}$ and $\mathbb{W}_e = \{W_{e1}, \cdots, W_{eq}\}$. In this paper, we call Model (2.1) the $\texttt{STAR}(p,q)$ model to reflect the fact that the data and the spatiotemporal weight matrices cover both space and time dimensions and to differentiate it from the commonly known $\texttt{SARAR}(p,q)$ model for pure cross-sectional data.[5] The exogenous regressors may include spatial Durbin terms, i.e., the spatiotemporal lags of (some) variables in X_N, but this would not affect the technical results.[6]

Considering the fact that the degree of dependence of the spatiotemporal observations will depend on at least the 'spatial distance' and 'time distance', it is evident that the use of spatiotemporal weight matrices (instead of spatial only) would be a useful modeling strategy for the type of spatiotemporal data considered in this paper. The type of spatial weight matrices may include within cluster interaction, between clusters interaction, economic connectivity matrix, physical connectivity matrix, *etc.*. The 'cluster' index in the table is essential in constructing the *cluster-membership* matrix. It can also be used to construct a *cluster interaction* spatial weight matrix. The transaction time together with the location of a transaction unit can be used to construct a *space-time* connectivity matrix. See Section 6 for details on the construction of the space-time connectivity matrix.

We first consider the cases where (a) the elements of ε_n are independent and identically distributed (*iid*), and (b) the elements of ε_n are independent but not identically distributed (*inid*). The number of clusters C and the number of time segments T can be: (i) both small, (ii) C small and T large, (iii) C large and T small, and (iv) both large.

In case of (a) and (i), the standard *quasi maximum likelihood* (QML) method can be followed for model estimation and inference. In all other cases, alternative methods are required due to the *incidental parameters problem* of Neyman and Scott (1948). We propose an *adjusted quasi score* (AQS) method for model estima-

[5]See Yang (2018) and the references therein for the theory and applications of $\texttt{SARAR}(p,q)$ model.

[6]The "spatiotemporal interaction" is captured by λ at mean level and by both λ and ρ at variance level.

tion and inference. As in the regular spatial econometrics models, QML and AQS methods will both face computational difficulties when general spatial weights matrices are used in the model, in terms of computing the determinants or inverse of these weights related matrices. In special cases where the space-time weights matrices are all lower triangular, this issue disappears (Pace et al. 1998a). In this paper, we propose a way of constructing these space-time connectivity matrices based on fixed *space window* and *time window*, so that the constructed matrices are 'band' matrices and hence the corresponding computational burden is alleviated.

2.3 QML Estimation of the High-Order STAR Model

Consider first the case that the number of clusters C and the number of time segments T are both fixed when sample size N grows, and the errors are independent and identically distributed (iid). In this case, we can merge the cluster and time effects into the regressors matrix X_N and our $\mathtt{STAR}$(p,q) model can be considered as a simple extension of the standard $\mathtt{SARAR}(p,q)$ model, the p-order spatial autoregressive ($\mathtt{SAR}$) model with q-order $\mathtt{SAR}$ disturbances. In this case, the QML method considered in Liu and Yang (2017) can be followed. To avoid the dummy variable trap, we delete the first column of M_T to give M_{T-1} and the first element of α to give α_{-1}. Letting $\mathbf{X}_N = [X_N, M_C, M_{T-1}]$ and $\boldsymbol{\beta} = (\beta', \mu', \alpha_2, \ldots, \alpha_T)'$, the model is written as

$$A_N(\lambda)Y_N = \mathbf{X}_N\boldsymbol{\beta} + V_N, \ \ B_N(\rho)V_N = \varepsilon_N, \tag{2.4}$$

where the elements of ε_N are assumed to be iid, and $\mathbf{X}_N$ is exogenously given.

The quasi Gaussian loglikelihood function for $\boldsymbol{\theta} = (\boldsymbol{\beta}', \sigma^2, \lambda', \rho')'$ is

$$\ell_N(\boldsymbol{\theta}) = -\frac{N}{2}\ln(2\pi\sigma^2) + \ln|A_N(\lambda)| + \ln|B_N(\rho)| - \frac{1}{2\sigma^2}\varepsilon'(\boldsymbol{\beta}, \delta)\varepsilon(\boldsymbol{\beta}, \delta), \tag{2.5}$$

where $\varepsilon_N(\boldsymbol{\beta}, \delta) = B_N(\rho)(A_N(\lambda)Y_N - \mathbf{X}_N\boldsymbol{\beta})$, and $|\cdot|$ denotes the determinant of a matrix. Maximizing (2.5) gives the maximum likelihood estimator (MLE) if the error term are indeed Gaussian, otherwise the quasi MLE (QMLE).

The maximization process can be simplified by first maximizing $\ell_N(\boldsymbol{\theta})$ for a given δ, to give the constrained QMLEs of $\boldsymbol{\beta}$ and σ^2:

$$\hat{\boldsymbol{\beta}}_N(\delta) = [\mathbb{X}'(\rho)\mathbb{X}(\rho)]^{-1}[\mathbb{X}'(\rho)\mathbb{Y}(\delta)], \tag{2.6}$$

$$\hat{\sigma}^2_N(\delta) = \tfrac{1}{N}\mathbb{Y}'(\delta)\mathbb{Q}(\rho)\mathbb{Y}(\delta), \tag{2.7}$$

where $\mathbb{Y}(\delta) = B_N(\rho)A_N(\lambda)Y_N$, $\mathbb{X}(\rho) = B_N(\rho)\mathbf{X}_N$, and $\mathbb{Q}(\rho) = I_N - \mathbb{X}(\rho)[\mathbb{X}(\rho)'\mathbb{X}(\rho)]^{-1}\mathbb{X}'(\rho)$. Then, substituting $\hat{\boldsymbol{\beta}}_N(\delta)$ and $\hat{\sigma}^2_N(\delta)$ back into (2.5) for $\boldsymbol{\beta}$ and σ^2, we obtain the concentrated loglikelihood function for δ:

$$\ell^c_N(\delta) = -\frac{N}{2}[\ln(2\pi)+1] - \frac{N}{2}\ln[\hat{\sigma}^2_N(\delta)] + \ln|A_N(\lambda)\,| + \ln|B_N(\rho)\,|\,. \tag{2.8}$$

Maximizing $\ell^c_N(\delta)$ gives the QMLE $\hat{\delta}_N$ of δ. Thus, the unconstrained QMLEs of $\boldsymbol{\beta}$ and σ^2 are $\hat{\boldsymbol{\beta}}_N \equiv \hat{\boldsymbol{\beta}}(\hat{\delta}_N)$ and $\hat{\sigma}^2_N \equiv \hat{\sigma}^2_N(\hat{\delta}_N)$, and that of $\boldsymbol{\theta}$ is $\hat{\boldsymbol{\theta}}_N = (\hat{\boldsymbol{\beta}}'_N, \hat{\sigma}^2_N, \hat{\delta}'_N)'$.

Under the regularity conditions (see Appendix A), $\hat{\boldsymbol{\theta}}_N$ is consistent for $\boldsymbol{\theta}$, and

$$\sqrt{N}(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}) \xrightarrow{D} N\big[0,\ \lim_{N\to\infty}\Sigma^{-1}_N(\boldsymbol{\theta})\Omega_N(\boldsymbol{\theta})\Sigma^{-1}_N(\boldsymbol{\theta})\big], \tag{2.9}$$

where $\Sigma_N(\boldsymbol{\theta}) = -\frac{1}{N}\mathrm{E}[\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\ell_N(\boldsymbol{\theta})]$ and $\Omega_N(\boldsymbol{\theta}) = \frac{1}{N}\mathrm{E}[S_N(\boldsymbol{\theta})S'_N(\boldsymbol{\theta})]$, with the limit of the former being a positive definite matrix, and that of the latter simply a constant matrix. The detailed expressions of these quantities are given below to facilitate practical applications.

First, it is useful to give a detailed expression of the quasi score (QS) vector, $S_N(\boldsymbol{\theta}) = \frac{\partial}{\partial\boldsymbol{\theta}}\ell_N(\boldsymbol{\theta})$, from which $\Sigma_N(\boldsymbol{\theta})$ and $\Omega_N(\boldsymbol{\theta})$ are derived (using Lemma B.3, Appendix B) and more importantly, the AQS estimation method to be considered in the next section is developed:

$$S_N(\boldsymbol{\theta}) = \begin{cases} \frac{1}{\sigma^2}\mathbb{X}'(\rho)\varepsilon(\boldsymbol{\beta},\delta) \\ \frac{1}{2\sigma^4}\varepsilon'_N(\boldsymbol{\beta},\delta)\varepsilon_N(\boldsymbol{\beta},\delta) - \frac{N}{2\sigma^2} \\ \frac{1}{\sigma^2}\varepsilon'_N(\boldsymbol{\beta},\delta)B_N(\rho)W_{\ell j}Y_N - \mathrm{tr}(F_{jN}(\lambda)), \qquad j = 1,\ldots,p, \\ \frac{1}{\sigma^2}\varepsilon'_N(\boldsymbol{\beta},\delta)G_{kN}(\rho)\varepsilon_N(\boldsymbol{\beta},\delta) - \mathrm{tr}(G_{kN}(\rho)), \quad k = 1,\ldots,q, \end{cases} \tag{2.10}$$

where $F_{jN}(\lambda) = W_{\ell j}A^{-1}_N(\lambda)$ and $G_{kN}(\rho) = W_{ek}B^{-1}_N(\rho)$, $j = 1,\ldots,p$ and $k = 1,\ldots,q$.

To give compact expressions for $\Sigma_n(\boldsymbol{\theta})$ and $\Omega_n(\boldsymbol{\theta})$, some notational conventions are followed: $\{a_j\}$ forms a row vector based on the elements $a_j,\ j = 1,\ldots,p$ (or q), $\{b_{jk}\}$ forms a matrix based on the elements $b_{jk}, j,k = 1,\ldots,p$ (or q), `diagv`$(\cdot)$

forms a column vector by the diagonal elements of a square matrix, and $A^s = A + A'$ for a matrix A. We have,

$$N\Sigma_N(\boldsymbol{\theta}) = \begin{bmatrix} \frac{1}{\sigma^2}\mathbb{X}'\mathbb{X} & 0 & \{\frac{1}{\sigma^2}\mathbb{X}'\eta_{jN}\} & 0 \\ \sim & \frac{N}{2\sigma^4} & \{\frac{1}{\sigma^2}\text{tr}(\bar{F}_{jN})\} & \{\frac{1}{\sigma^2}\text{tr}(G_{kN})\} \\ \sim & \sim & \{\frac{1}{\sigma^2}\eta'_{jN}\eta_{j'N} + \text{tr}(\bar{F}^s_{jN}\bar{F}_{j'N})\} & \{\text{tr}(G^s_{kN}\bar{F}_{jN})\} \\ \sim & \sim & \sim & \{\text{tr}(G^s_{kN}G_{k'N})\} \end{bmatrix},$$

where $\mathbb{X} = \mathbb{X}(\rho)$, $F_{jN} = F_{jN}(\lambda)$, $G_{kN} = G_{kN}(\rho)$, $\bar{F}_{jN} \equiv \bar{F}_{jN}(\delta) = B_N(\rho)F_{jN}(\lambda)B_N^{-1}(\rho)$, and $\eta_{jN} \equiv \eta_{jN}(\boldsymbol{\beta}, \delta) = B_N(\rho)F_{jN}(\lambda)\mathbf{X}_N\boldsymbol{\beta}$; and $\Omega_N(\boldsymbol{\theta}) = \Sigma_N(\boldsymbol{\theta}) + \Gamma_N(\boldsymbol{\theta})$, where

$$N\Gamma_N(\boldsymbol{\theta}) = \begin{bmatrix} 0_{K\times K} & \frac{\gamma}{2\sigma^3}\mathbb{X}'\iota_N & \{\frac{\gamma}{\sigma}\mathbb{X}'\bar{f}_{jN}\} & \{\frac{\gamma}{\sigma}\mathbb{X}'g_{kN}\} \\ \sim & \frac{N\kappa}{4\sigma^4} & \{\frac{\kappa}{2\sigma^2}\text{tr}(\bar{F}_{jN}) + \frac{\gamma}{2\sigma^3}\iota'_N\eta_{jN}\} & \{\frac{\kappa}{2\sigma^2}\text{tr}(G_{kN})\} \\ \sim & \sim & \{\kappa\bar{f}'_{jN}\bar{f}_{j'N} + \frac{\gamma}{\sigma}\bar{f}'_{jN}\eta_{j'N} + \frac{\gamma}{\sigma}\bar{f}'_{j'N}\eta_{jN}\} & \{\kappa\bar{f}'_{jN}g_{kN} + \frac{\gamma}{\sigma}\eta'_{jN}g_{kN}\} \\ \sim & \sim & \sim & \{\kappa g'_{kN}g_{k'N}\} \end{bmatrix},$$

$\gamma =$ skewness of ε_{ni}, $\kappa =$ excess kurtosis of ε_{ni}, $\bar{f}_{jN} = \texttt{diagv}(\bar{F}_{jN})$ and $g_{kN} = \texttt{diagv}(G_{kN})$.

In real applications, $\Sigma_N(\boldsymbol{\theta})$ is estimated by $\Sigma_N(\hat{\boldsymbol{\theta}})$, and $\Omega_N(\boldsymbol{\theta})$ by $\Omega_N(\hat{\boldsymbol{\theta}})$. γ and κ are estimated by the sample skewness and excess kurtosis of the QML residuals $\{\hat{\varepsilon}_{ni}\}$.

2.4 AQS Estimation of the High-Order STAR Model

As sample size N increases, C or T or both may increase with N. As a result, the number of parameters (in μ or α or both) increases with N, giving rise to the *incidental parameters problem* of Neyman and Scott (1948)[7]. This makes the QML method invalid. As the type of data we consider is not panel data, the standard transformation method for standard spatial panel data (like Lee and Yu (2010a)) cannot be applied to wipe out the fixed effects. To overcome this difficulty, we propose an *adjusted quasi score* (AQS) method.

Denote $\phi = (\mu', \alpha_2, \ldots, \alpha_T)'$. Partition $\mathbb{X}(\rho)$ defined below (2.7) into $[\mathbb{X}_1(\rho),\ \mathbb{X}_2(\rho)] \equiv \{B_N(\rho)X_N,\ B_N(\rho)[M_C, M_{T-1}]\}$. Given β and δ, $\ell_N(\boldsymbol{\theta})$ given in (2.5) is maximized at

[7] To have a clear view of the history of *incidental parameters problem*, see Lancaster (2000).

$$\hat{\phi}_N(\beta,\delta) = [\mathbb{X}_2'(\rho)\mathbb{X}_2(\rho)]^{-1}\mathbb{X}_2'(\rho)[\mathbb{Y}(\delta) - \mathbb{X}_1(\rho)\beta], \tag{2.11}$$

where $\mathbb{Y}(\delta)$ is defined below (2.7). Substituting $\hat{\phi}_N(\theta)$ back into $\ell_N(\boldsymbol{\theta})$ for ϕ, we obtain the concentrated loglikelihood function of $\theta = (\beta', \sigma^2, \delta')'$:

$$\ell_N^c(\theta) = -\frac{N}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\tilde{\varepsilon}_N'(\beta,\delta)\tilde{\varepsilon}_N(\beta,\delta) + \ln|A_N(\lambda)| + \ln|B_N(\rho)|, \tag{2.12}$$

where $\tilde{\varepsilon}_N(\beta,\delta) = \mathbb{Q}_2(\rho)[\mathbb{Y}(\delta) - \mathbb{X}_1(\rho)\beta]$ and $\mathbb{Q}_2(\rho) = I_N - \mathbb{X}_2(\rho)[\mathbb{X}_2'(\rho)\mathbb{X}_2(\rho)]^{-1}\mathbb{X}_2'(\rho)$.

Taking the partial derivatives of $\ell_N^c(\theta)$ gives the concentrated quasi score function of θ. Or equivalently, solve $S_{N,\phi}(\boldsymbol{\theta}) = 0$ to give $\hat{\phi}_N(\theta)$ and then substitute $\hat{\phi}_N(\theta)$ into $S_{N,\theta}(\boldsymbol{\theta})$, where $S_{N,\phi}(\boldsymbol{\theta})$ and $S_{N,\theta}(\boldsymbol{\theta})$ denote, respectively, the ϕ- and θ-component of the QS function $S_N(\boldsymbol{\theta})$ given in (2.10). The concentrated QS (CQS) function for θ takes the form:

$$S_N^c(\theta) = \begin{cases} \frac{1}{\sigma^2}\mathbb{X}_1'(\rho)\tilde{\varepsilon}_N(\beta,\delta), \\ \frac{1}{2\sigma^4}\tilde{\varepsilon}_N'(\beta,\delta)\tilde{\varepsilon}_N(\beta,\delta) - \frac{N}{2\sigma^2}, \\ \frac{1}{\sigma^2}\tilde{\varepsilon}_N'(\beta,\delta)B_N(\rho)W_{\ell j}Y_N - \mathrm{tr}(F_{jN}(\lambda)), \quad j = 1,\ldots,p, \\ \frac{1}{\sigma^2}\tilde{\varepsilon}_N'(\beta,\delta)G_{kN}(\rho)\tilde{\varepsilon}_N(\beta,\delta) - \mathrm{tr}(G_{kN}(\rho)), \ k = 1,\ldots,q, \end{cases} \tag{2.13}$$

and its expectation at θ (when it represents the true parameter values):

$$\mathrm{E}[S_N^c(\theta)] = \begin{cases} 0, \\ \frac{N-(C+T-1)}{2\sigma^2} - \frac{N}{2\sigma^2}, \\ \mathrm{tr}(\mathbb{Q}_2(\rho)\bar{F}_{jN}(\delta)) - \mathrm{tr}(F_{jN}(\lambda)), \quad j = 1,\ldots,p, \\ \mathrm{tr}(\mathbb{Q}_2(\rho)G_{kN}(\rho)) - \mathrm{tr}(G_{kN}(\rho)), \ k = 1,\ldots,q, \end{cases} \tag{2.14}$$

where $F_{jN}(\lambda)$, $\bar{F}_{jN}(\lambda)$ and $G_{kN}(\rho)$ are defined in Sec. 2.3.

Therefore, $\mathrm{E}[S_N^c(\theta)] \neq 0$ and hence the regular QMLE will incur bias. Furthermore, when C or T grows with N, it can be that $\lim_{N\to\infty}\frac{1}{N}\mathrm{E}[S_N^c(\theta)] \neq 0$ and that $\mathrm{plim}_{N\to\infty}\frac{1}{N}S_N^c(\theta) \neq 0$. Subsequently, the QMLE will not be consistent (the *incidental parameters problem*). To solve this problem, we adjust the CQS function so that it satisfies the necessary conditions for a consistent estimation. Define the AQS function as $S_N^*(\theta) = S_N^c(\theta) - \mathrm{E}[S_N^c(\theta)]$, i.e.,

$$S_N^*(\theta) = \begin{cases} \frac{1}{\sigma^2}\mathbb{X}_1'(\rho)\tilde{\varepsilon}_N(\beta,\delta) \\ \frac{1}{2\sigma^4}\tilde{\varepsilon}_N'(\beta,\delta)\tilde{\varepsilon}_N(\beta,\delta) - \frac{N-(C+T-1)}{2\sigma^2} \\ \frac{1}{\sigma^2}\tilde{\varepsilon}_N'(\beta,\delta)B_N(\rho)W_{\ell j}Y_N - \mathrm{tr}(\mathbb{Q}_2(\rho)\bar{F}_{jN}(\delta)), \quad j = 1,\dots,p, \\ \frac{1}{\sigma^2}\tilde{\varepsilon}_N'(\beta,\delta)G_{kN}(\rho)\tilde{\varepsilon}_N(\beta,\delta) - \mathrm{tr}(\mathbb{Q}_2(\rho)G_{kN}(\rho)), \;\; k = 1,\dots,q. \end{cases} \tag{2.15}$$

Correcting the concentrated quasi scores removes the effects of estimating the *incidental parameters*, which are the cluster and time effects. Solving $S_N^*(\theta) = 0$ gives the AQS estimator $\hat{\theta}_{\mathtt{AQS}}$ of the common parameters θ. It can be shown that under mild regularity conditions (see Appendix A), $\hat{\theta}_{\mathtt{AQS}} - \theta \xrightarrow{p} 0$, and

$$\sqrt{N}(\hat{\theta}_{\mathtt{AQS}} - \theta) \xrightarrow{D} N\big[0,\ \lim_{N\to\infty}\Sigma_N^{*-1}(\theta)\Omega_N^*(\theta)\Sigma_N^{*\prime-1}(\theta)\big], \tag{2.16}$$

where $\Sigma_N^*(\theta) = -\frac{1}{N}\mathrm{E}[\frac{\partial}{\partial\theta'}S_N^*(\theta)]$ and $\Omega_N^*(\theta) = \frac{1}{N}\mathrm{E}[S_N^*(\theta)S_N^{*\prime}(\theta)]$, with the former being assumed to be invertible for large enough N, and the latter being assumed to exist. Note that $\Sigma_N^*(\theta)$ is asymmetric in λ and ρ elements due to the adjustments on the CQS functions.

The process of finding the root of $S_N^*(\theta) = 0$ can be simplified by first solving the first two components for β and σ^2 to give

$$\hat{\beta}_{\mathtt{AQS}}(\delta) = [\mathbb{X}_1'(\rho)\mathbb{Q}_2(\rho)\mathbb{X}_1(\rho)]^{-1}\mathbb{X}_1'(\rho)\mathbb{Q}_2(\rho)\mathbb{Y}(\delta), \tag{2.17}$$

$$\hat{\sigma}^2_{\mathtt{AQS}}(\delta) = \frac{1}{N-(C+T-1)}\,[\mathbb{Y}(\delta) - \mathbb{X}_1(\rho)\hat{\beta}_{\mathtt{AQS}}(\delta)]'\mathbb{Q}_2(\rho)[\mathbb{Y}(\delta) - \mathbb{X}_1(\rho)\hat{\beta}_{\mathtt{AQS}}(\delta)], \tag{2.18}$$

and then substituting $\hat{\beta}_{\mathtt{AQS}}(\delta)$ and $\hat{\sigma}^2_{\mathtt{AQS}}(\delta)$ back into the last two sets of equations for λ and ρ to give the concentrated AQS function:

$$S_N^{*c}(\delta) = \begin{cases} \frac{1}{\hat{\sigma}^2_{\mathtt{AQS}}(\delta)}\tilde{\varepsilon}_N'(\hat{\beta}_{\mathtt{AQS}}(\delta),\delta)B_N(\rho)W_{\ell j}Y_N - \mathrm{tr}(\mathbb{Q}_2(\rho)\bar{F}_{jN}(\delta)), \quad j = 1,\dots,p, \\ \frac{1}{\hat{\sigma}^2_{\mathtt{AQS}}(\delta)}\tilde{\varepsilon}_N'(\hat{\beta}_{\mathtt{AQS}}(\delta),\delta)G_{kN}(\rho)\tilde{\varepsilon}_N(\hat{\beta}_{\mathtt{AQS}}(\delta),\delta) - \mathrm{tr}(\mathbb{Q}_2(\rho)G_{kN}(\rho)), \;\; k = 1,\dots,q. \end{cases} \tag{2.19}$$

Solving $S_N^{*c}(\delta) = 0$ gives the AQS estimator $\hat{\delta}_{\mathtt{AQS}}$ of δ, which in turn gives the AQS estimators $\hat{\beta}_{\mathtt{AQS}} = \hat{\beta}_{\mathtt{AQS}}(\hat{\delta}_{\mathtt{AQS}})$ and $\hat{\sigma}^2_{\mathtt{AQS}} = \hat{\sigma}^2_{\mathtt{AQS}}(\hat{\delta}_{\mathtt{AQS}})$ of β and σ^2. Thus, $\hat{\theta}_{\mathtt{AQS}} = (\hat{\beta}_{\mathtt{AQS}}', \hat{\sigma}^2_{\mathtt{AQS}}, \hat{\delta}_{\mathtt{AQS}}')'$.

For conducting statistical inferences, we need to provide methods of estimating the asymptotic variance of $\hat{\theta}_{\texttt{AQS}}$. The matrix $\Sigma_N^*(\theta)$ can be consistently estimated by $\frac{1}{N}H_N^*(\hat{\theta}_{\texttt{AQS}}^*)$, where $H_N^*(\theta) = -\frac{\partial}{\partial\theta'}S_N^*(\theta)$ is the negative Hessian matrix. To derive $H_N^*(\theta)$, note that $\tilde{\varepsilon}_N(\beta,\delta) = \mathbb{Q}_2(\rho)[\mathbb{Y}(\delta) - \mathbb{X}_1(\rho)\beta]$ and that $\frac{\partial}{\partial\rho_k}\mathbb{Q}_2(\rho) = \mathbb{Q}_2(\rho)G_{kN}(\rho)\mathbb{P}_2(\rho) + \mathbb{P}_2(\rho)G'_{kN}(\rho)\mathbb{Q}_2(\rho)$, where $\mathbb{P}_2(\rho) = I_N - \mathbb{Q}_2(\rho)$. Following the notational conventions introduced in Section 2.3 and using again the short-hand notations, e.g., $\mathbb{X}_1 = \mathbb{X}_1(\rho)$, $\mathbb{Q}_2 = \mathbb{Q}_2(\rho)$, $\tilde{\varepsilon}_N = \tilde{\varepsilon}_N(\beta,\delta)$, the elements of $H_N^*(\theta)$ are written as follows:

$$
\begin{aligned}
H^*_{\beta\theta}(\theta) &= \left[\tfrac{1}{\sigma^2}\mathbb{X}_1'\mathbb{Q}_2\mathbb{X}_1,\ \tfrac{1}{\sigma^4}\mathbb{X}_1'\tilde{\varepsilon}_N,\ \left\{\tfrac{1}{\sigma^2}\mathbb{X}_1'\mathbb{Q}_2\bar{F}_{jN}\mathbb{Y}\right\},\ \left\{\tfrac{1}{\sigma^2}\mathbb{X}_1'\mathbb{Q}_2G^s_{kN}\tilde{\varepsilon}_N\right\}\right],\\
H^*_{\sigma^2\theta}(\theta) &= \left[\tfrac{1}{\sigma^4}\tilde{\varepsilon}_N'\mathbb{X}_1,\ \tfrac{1}{\sigma^6}\tilde{\varepsilon}_N'\tilde{\varepsilon}_N - \tfrac{N-(C+T-1)}{2\sigma^4},\ \left\{\tfrac{1}{\sigma^4}\tilde{\varepsilon}_N'\bar{F}_{jN}\mathbb{Y}\right\},\ \left\{\tfrac{1}{\sigma^4}\tilde{\varepsilon}_N'G_{kN}\tilde{\varepsilon}_N\right\}\right],\\
H^*_{\lambda\beta}(\theta) &= \left\{\tfrac{1}{\sigma^2}\mathbb{X}_1'\mathbb{Q}_2\bar{F}_{jN}\mathbb{Y}\right\}', \quad H^*_{\lambda\sigma^2}(\theta) = \left\{\tfrac{1}{\sigma^4}\tilde{\varepsilon}_N'\bar{F}_{jN}\mathbb{Y}\right\}',\\
H^*_{\lambda\lambda}(\theta) &= \left\{\tfrac{1}{\sigma^2}\mathbb{Y}'\bar{F}_{jN}'\mathbb{Q}_2\bar{F}_{j'N}\mathbb{Y} + \operatorname{tr}(\mathbb{Q}_2\bar{F}_{jN}\bar{F}_{j'N})\right\},\\
H^*_{\lambda\rho}(\theta) &= \left\{\tfrac{1}{\sigma^2}\tilde{\varepsilon}_N'G^s_{kN}\mathbb{Q}_2\bar{F}_{jN}\mathbb{Y} + \operatorname{tr}(\mathbb{P}_2G^s_{kN}\mathbb{Q}_2\bar{F}_{jN})\right\},\\
H^*_{\rho\lambda}(\theta) &= \left\{\tfrac{1}{\sigma^2}\tilde{\varepsilon}_N'G^s_{kN}\mathbb{Q}_2\bar{F}_{jN}\mathbb{Y}\right\}',\ H^*_{\rho\beta}(\theta) = H^{*\prime}_{\beta\rho}(\theta),\ H^*_{\rho\sigma^2}(\theta) = H^{*\prime}_{\sigma^2\rho}(\theta),\\
H^*_{\rho\rho}(\theta) &= \left\{\tfrac{1}{\sigma^2}\tilde{\varepsilon}_N'(G'_{kN}G_{k'N} - G^s_{kN}\mathbb{P}_2G^s_{k'N})\tilde{\varepsilon}_N + \operatorname{tr}[\mathbb{Q}_2(G_{kN}\mathbb{P}_2G^s_{k'N} + G_{k'N}G_{kN})]\right\}.
\end{aligned}
$$

From the expression of $H_N^*(\theta)$, the analytical expression of $\Sigma_N^*(\theta)$ can be found but it is not necessary to do so as $H_N^*(\hat{\theta}_{\texttt{AQS}})$ provides a consistent estimator of $\Sigma_N^*(\theta)$ in the sense that $\frac{1}{N}[H_N^*(\hat{\theta}_{\texttt{AQS}}) - \Sigma_N^*(\theta)] = o_p(1)$. Besides, $\Sigma_N^*(\theta)$ involves ϕ, which may not be consistently estimated when either C or T goes large with N proportionally.

As $\Omega_N^*(\theta)$ does not have a simple sample analogue as does $\Sigma_N^*(\theta)$, one may reply on its analytical expression for a plug-in type estimation. The $\Omega_N^*(\theta)$ matrix, derived using Lemma B.3 (Appendix B), has the distinct elements:

$$
\begin{aligned}
N\Omega^*_{\beta\theta}(\theta) &= \left[\tfrac{1}{\sigma^2}\mathbb{X}_1'\mathbb{Q}_2\mathbb{X}_1,\ \tfrac{\gamma}{2\sigma^3}\mathbb{X}_1'\mathbb{Q}_2q_2,\ \left\{\tfrac{1}{\sigma^2}\mathbb{X}_1'\mathbb{Q}_2\eta_{jN} + \tfrac{\gamma}{\sigma}\mathbb{X}_1'\mathbb{Q}_2\bar{f}_{jN}\right\},\ \left\{\tfrac{\gamma}{\sigma}\mathbb{X}_1'\mathbb{Q}_2\bar{g}_{kN}\right\}\right],\\
N\Omega^*_{\sigma^2\sigma^2}(\theta) &= \tfrac{\kappa}{4\sigma^4}q_2'q_2 + \tfrac{N-(C+T-1)}{2\sigma^4},
\end{aligned}
$$

$$
\begin{aligned}
N\Omega^*_{\sigma^2\lambda}(\theta) &= \left\{\tfrac{\gamma}{2\sigma^3} q_2' \mathbb{Q}_2 \eta_{jN} + \tfrac{\kappa}{2\sigma^2} q_2' \bar{f}_{jN} + \tfrac{1}{\sigma^2}\mathrm{tr}(\mathbb{Q}_2 \bar{F}_{jN})\right\},\\
N\Omega^*_{\sigma^2\rho}(\theta) &= \left\{\tfrac{\kappa}{2\sigma^2} q_2' \bar{g}_{kN} + \tfrac{1}{\sigma^2}\mathrm{tr}(\mathbb{Q}_2 G_{kN})\right\},\\
N\Omega^*_{\lambda\lambda}(\theta) &= \left\{\tfrac{1}{\sigma^2}\eta_{jN}'\mathbb{Q}_2\eta_{j'N} + \tfrac{\gamma}{\sigma}\eta_{jN}'\mathbb{Q}_2\bar{f}_{j'N} + \tfrac{\gamma}{\sigma}\eta_{j'N}'\mathbb{Q}_2\bar{f}_{jN} + \kappa\bar{f}_{jN}'\bar{f}_{j'N}\right.\\
&\qquad \left. + \mathrm{tr}(\mathbb{Q}_2\bar{F}_{jN}(\mathbb{Q}_2\bar{F}_{j'N} + \bar{F}_{j'N}'\mathbb{Q}_2))\right\},\\
N\Omega^*_{\lambda\rho}(\theta) &= \left\{\tfrac{\gamma}{\sigma}\eta_{jN}'\mathbb{Q}_2\bar{g}_{kN} + \kappa\bar{f}_{jN}'\bar{g}_{kN} + \mathrm{tr}(\bar{F}_{jN}\mathbb{Q}_2 G^s_{kN}\mathbb{Q}_2)\right\},\\
N\Omega^*_{\rho\rho}(\theta) &= \left\{\kappa\bar{g}_{kN}'\bar{g}_{k'N} + \mathrm{tr}(G_{kN}\mathbb{Q}_2 G^s_{k'N}\mathbb{Q}_2)\right\},
\end{aligned}
$$

where γ and κ are defined in Section 2.3, $\eta_{jN} \equiv \eta_{jN}(\beta,\delta) = \bar{F}_{jN}(\delta)[\mathbb{X}_1(\rho)\beta + \mathbb{X}_2(\rho)\phi]$, $\mathbb{Q}_2 = \mathbb{Q}_2(\rho)$, $q_2 = \mathtt{diagv}(\mathbb{Q}_2)$, $\bar{f}_{jN} = \mathtt{diagv}(\mathbb{Q}_2\bar{F}_{jN})$, $\bar{g}_{kN} = \mathtt{diagv}(\mathbb{Q}_2 G_{kN}\mathbb{Q}_2)$.

From its analytical expression, we see that $\Omega^*_N(\theta)$ contains additional parameters, γ, κ and ϕ (through η_N), besides the common parameters of interest θ. The plug-in estimator of $\Omega^*_N(\theta)$ therefore involves two issues: one is the consistent estimation of γ and κ and the other is the effect of plugging in $\hat{\phi}_{\mathtt{AQS}}$. To address first issue, note that the original errors $\varepsilon_N = \mathbb{Y} - \mathbb{X}\boldsymbol{\beta}$ may not be consistently estimated due to its involvement of ϕ, but the transformed errors $\tilde{\varepsilon}_N = \mathbb{Q}_2(\mathbb{Y} - \mathbb{X}\boldsymbol{\beta}) = \mathbb{Q}_2\varepsilon_N$ can be consistently estimated in general by its AQS counterpart, $\hat{\varepsilon}_{\mathtt{AQS}} = \mathbb{Q}_2(\hat{\rho}_{\mathtt{AQS}})[\mathbb{Y}(\hat{\delta}_{\mathtt{AQS}}) - \mathbb{X}_1(\hat{\rho}_{\mathtt{AQS}})\hat{\beta}_{\mathtt{AQS}}]$. Let $q_{2,ij}$ be the $(i,\ j)$th element of $\mathbb{Q}_2$. Since

$$\tilde{v}_i = q_{2,i1}v_1 + q_{2,i2}v_2 + \cdots + q_{2,iN}v_N,$$

where $\tilde{v}_i$ and v_i are the elements in $\tilde{\varepsilon}_N$ and ε_N. As

$$\mathrm{E}(\tilde{v}_i^3) = \textstyle\sum_{j=1}^N q_{2,ij}^3 \mathrm{E}\left(v_j^3\right) = \sigma^3\gamma\sum_{j=1}^N q_{2,ij}^3,$$

we then have the consistent estimator for γ, which is given by

$$\hat{\gamma}_{\mathtt{AQS}} = \frac{\sum_{i=1}^N \hat{v}_{\mathtt{AQS},i}^3}{\hat{\sigma}_{\mathtt{AQS}}^3\sum_{i=1}^N\sum_{j=1}^N \hat{q}_{2,ij}^3}.$$

Similarly, as

$$
\begin{aligned}
\mathrm{E}(\tilde{v}_i^4) &= \textstyle\sum_{j=1}^N q_{2,ij}^4\mathrm{E}(v_j^4) + 3\sigma^4\sum_{j=1}^N\sum_{l=1}^N q_{2,ij}^2 q_{2,il}^2 - 3\sigma^4\sum_{j=1}^N q_{2,ij}^4\\
&= \textstyle\sum_{j=1}^N q_{2,ij}^4\kappa\sigma^4 + 3\sigma^4\sum_{j=1}^N\sum_{l=1}^N q_{2,ij}^2 q_{2,il}^2,
\end{aligned}
$$

we have the consistent estimator for κ, which is given by

$$\hat{\kappa}_{\texttt{AQS}} = \frac{\sum_{i=1}^{N} \hat{v}_{\texttt{AQS},i}^{4} - 3\hat{\sigma}_{\texttt{AQS}}^{4} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{l=1}^{N} \hat{q}_{2,ij}^{2} \hat{q}_{2,il}^{2}}{\hat{\sigma}_{\texttt{AQS}}^{4} \sum_{i=1}^{N} \sum_{j=1}^{N} \hat{q}_{2,ij}^{4}}$$

Thus, we obtain a plug-in estimator $\widehat{\Omega}_N^*$ by plugging $\hat{\theta}_{\texttt{AQS}}$, $\hat{\gamma}_{\texttt{AQS}}$, $\hat{\kappa}_{\texttt{AQS}}$, and $\hat{\phi}_{\texttt{AQS}}$ into $\Omega_N^*(\theta)$.

The plug-in estimator $\widehat{\Omega}_N^*$ is valid when neither C nor T grows with N proportionally, but may not be valid if either C or T grows with N proportionally, as in this case consistent estimation of ϕ may not be achieved.[8] To address the second issue facing the plug-in estimator $\widehat{\Omega}_N^*$, note that ϕ appears in $\Omega_N(\theta)$ either linearly or quadratically through η_{jN}. To see the effect of replacing ϕ by $\hat{\phi}_{\texttt{AQS}}$ in $\Omega_N(\theta)$, we note from (2.11) that (dropping the subscript $_{\texttt{AQS}}$),

$$\hat{\phi} = [\mathbb{X}_2'(\hat{\rho})\mathbb{X}_2(\hat{\rho})]^{-1}\mathbb{X}_2'(\hat{\rho})[\mathbb{Y}(\hat{\delta}) - \mathbb{X}_1(\hat{\rho})\hat{\beta}] = \phi + (\mathbb{X}_2'\mathbb{X}_2)^{-1}\mathbb{X}_2'\varepsilon_N, \qquad (2.20)$$

by the consistency of $\hat{\beta}, \hat{\rho}$ and $\hat{\delta}$. Based on this, it is not difficult to show that

$$\hat{\mu}_c = \mu_c + o_p(\tfrac{1}{\sqrt{n_c}}) \quad \text{and} \quad \hat{\alpha}_t = \alpha_t + o_p(\tfrac{1}{\sqrt{m_t}}),$$

where μ_c is the cth cluster effect, n_c is the number of transactions belong to cluster c, α_t is the tth period effect, and m_t is the number of transactions occurred in period t. Therefore, when $\{n_c\}$ are fixed, meaning that C grows with N proportionally, the estimators $\{\hat{\mu}_c\}$ are not consistent. Similarly, when $\{m_t\}$ are fixed, meaning that T grows with N proportionally, the estimators $\{\hat{\alpha}_t\}$ are not consistent.

From (2.20), one can easily see that for the terms linear in ϕ, the effect of replacing ϕ by $\hat{\phi}_{\texttt{AQS}}$ is asymptotically negligible. However, for the sole quadratic in ϕ term, $\phi'\mathbb{X}_2'\bar{F}_{jN}'\mathbb{Q}_2\bar{F}_{j'N}\mathbb{X}_2\phi$, which is embedded in $\eta_{jN}'\mathbb{Q}_2\eta_{j'N}$, its plug-in estimator

[8] As ϕ is a set of incidental parameters whose dimension grows with N, its AQS estimator $\hat{\phi}_{\texttt{AQS}} = \hat{\phi}_N(\hat{\beta}_{\texttt{AQS}}, \hat{\delta}_{\texttt{AQS}})$ obtained through (2.11) may not be consistent, unless as N increases information on each element of ϕ accumulates, i.e., both the cluster sizes n_c and the number of transactions m_t occurred in period t increase with N. When C increases but $\{n_c\}$ are fixed, the information on each μ-element does not accumulate and $\hat{\mu}_{\texttt{AQS}}$ is not consistent. Similarly, if T increases but $\{m_t\}$ are fixed, $\hat{\alpha}_{\texttt{AQS}}$ is not consistent.

is seen to be,

$$
\begin{aligned}
&\hat{\phi}'\mathbb{X}_2'(\hat{\rho})\bar{F}_{jN}'(\hat{\delta})\mathbb{Q}_2\bar{F}_{j'N}(\hat{\delta})\mathbb{X}_2(\hat{\rho})\hat{\phi} \\
&= \phi'\mathbb{X}_2'\bar{F}_{jN}'\mathbb{Q}_2\bar{F}_{j'N}\mathbb{X}_2\phi + \varepsilon_N'\mathbb{P}_2\bar{F}_{jN}'\mathbb{Q}_2\bar{F}_{j'N}\mathbb{P}_2\varepsilon_N + o_p(N) \\
&= \phi'\mathbb{X}_2'\bar{F}_{jN}'\mathbb{Q}_2\bar{F}_{j'N}\mathbb{X}_2\phi + \sigma^2\mathrm{tr}(\mathbb{P}_2\bar{F}_{jN}'\mathbb{Q}_2\bar{F}_{j'N}) + o_p(N). \qquad (2.21)
\end{aligned}
$$

Therefore, the plug-in estimator of the term $\frac{1}{\sigma^2}\eta_{jN}'\mathbb{Q}_2\eta_{j'N}$ in $N\Omega_N^*(\theta)$ should be bias-corrected by subtracting the plug-in estimator of $\mathrm{tr}(\mathbb{P}_2\bar{F}_{jN}'\mathbb{Q}_2\bar{F}_{j'N})$.

2.5 AQS Estimation under Heteroskedasticity

Both the QML and AQS methods considered in the early sections are based on the assumption that the errors $\{\varepsilon_i\}$ are iid. The iid assumption is often questionable for spatial data, and the inid assumption or *unknown heteroskedasticity* may be more realistic in practical applications, especially for the real estate prices data. In this case, both the methods considered above are invalid. The unknown heteroskedasticity (UH) brings in another set of incidental parameters. To overcome this difficulty, we develop an AQS method that is robust against UH. The key idea for achieving this is to adjust the quasi score functions so that thir expectations at the true parameters remain zero under UH.

For ease of exposition, additional notational conventions are followed: $\mathtt{diag}(\cdot)$ forms a diagonal matrix based on diagonal elements of a square matrix or a vector, to differentiate it from $\mathtt{diagv}(\cdot)$ introduced earlier, and $A^\circ = A - \mathtt{diag}(A)$ for a square matrix A.

Assume that $\varepsilon_N \sim (0, H_N)$, where $H_N = \mathrm{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_N^2)$, σ_i^2 is the variance of the error ε_i. In this section, we do not include the σ^2 element of the score vector as we cannot do inference for each σ_i^2, though we can consistently estimate the average of the error variance.

2.5.1 Case of fixed C and T

Consider first the case where C and T are both fixed as N increases. Under heteroskedasticity, the quasi score function $S_N(\boldsymbol{\theta})$ given in (2.10) no longer has a zero expectation at the true parameters, and the necessary condition for consistent estimation, $\text{plim}_{N\to\infty}\frac{1}{N}S_N(\boldsymbol{\theta}) = 0$, may not be satisfied. However, we notice that its $\boldsymbol{\beta}$ component continues to have a zero expectation, and hence does not need to be further adjusted. Write the stochastic parts of the λ-components of $S_N(\boldsymbol{\theta})$ as $\varepsilon_N'(\boldsymbol{\beta},\delta)B_N(\rho)W_{\ell j}Y_N = \varepsilon_N'(\boldsymbol{\beta},\delta)\bar{F}_{jN}(\delta)\mathbb{Y}(\delta), j = 1,\ldots,p$. Simply replacing $\bar{F}_{jN}(\delta)$ by $\bar{F}_{jN}^{\circ}(\delta)$ gives a set of estimating functions for λ that has the desired property. Similarly, replacing $G_{kN}(\rho)$ by $G_{kN}^{\circ}(\rho)$ in the ρ-components of $S_N(\boldsymbol{\theta})$ gives a set of desired estimating functions for ρ. Thus, the set of AQS functions for $\boldsymbol{\vartheta} = (\boldsymbol{\beta}',\lambda',\rho')'$ robust against unknown heteroskedasticity is simply:

$$S_N^{\dagger}(\boldsymbol{\vartheta}) = \begin{cases} \mathbb{X}'(\rho)\varepsilon(\boldsymbol{\beta},\delta) \\ \varepsilon_N'(\boldsymbol{\beta},\delta)\bar{F}_{jN}^{\circ}(\delta)\mathbb{Y}(\delta), \quad j = 1,\ldots,p, \\ \varepsilon_N'(\boldsymbol{\beta},\delta)G_{kN}^{\circ}(\rho)\varepsilon_N(\boldsymbol{\beta},\delta), \quad k = 1,\ldots,q. \end{cases} \tag{2.22}$$

Similar to (2.6) and (2.7) which partially solve the QS equations formed by (2.10), to find the root of the robust AQS functions, we first solve the $\boldsymbol{\beta}$ component of $S_N^{\dagger}(\boldsymbol{\vartheta}) = 0$ to give,

$$\hat{\boldsymbol{\beta}}_{\mathrm{R}}(\delta) = [\mathbb{X}'(\rho)\mathbb{X}(\rho)]^{-1}[\mathbb{X}'(\rho)\mathbb{Y}(\delta)], \tag{2.23}$$

and then substituting $\hat{\boldsymbol{\beta}}_{\mathrm{R}}(\delta)$ back into the last two robust AQS functions in (2.22) to give the concentrated robust AQS function:

$$S_N^{\dagger c}(\delta) = \begin{cases} \varepsilon_N'(\hat{\boldsymbol{\beta}}_{\mathrm{R}}(\delta),\delta)\bar{F}_{jN}^{\circ}(\delta)\mathbb{Y}(\delta), \quad j = 1,\ldots,p, \\ \varepsilon_N'(\hat{\boldsymbol{\beta}}_{\mathrm{R}}(\delta),\delta)G_{kN}^{\circ}(\delta)\varepsilon_N(\hat{\boldsymbol{\beta}}_{\mathrm{R}}(\delta),\delta), \ k = 1,\ldots,q. \end{cases} \tag{2.24}$$

Solving $S_N^{\dagger c}(\delta) = 0$ gives the robust AQS estimator $\hat{\delta}_{\mathrm{R}}$ of δ. Plugging it back into (2.23) gives the AQS estimator $\hat{\boldsymbol{\beta}}_{\mathrm{R}} = \hat{\boldsymbol{\beta}}_{\mathrm{R}}(\hat{\delta}_{\mathrm{R}})$ of β, and thus, a consistent estimator $\hat{\boldsymbol{\vartheta}}_{\mathrm{R}} = (\hat{\beta}_{\mathrm{R}}', \hat{\delta}_{\mathrm{R}}')'$ for $\boldsymbol{\vartheta}$.

It can be shown that under mild regularity conditions (in Appendix A) $\hat{\boldsymbol{\vartheta}}_{\mathrm{R}} - \boldsymbol{\vartheta} \xrightarrow{p} 0$, and

$$\sqrt{N}(\hat{\boldsymbol{\vartheta}}_{\mathrm{R}} - \boldsymbol{\vartheta}) \xrightarrow{D} N\big[0, \ \lim_{N\to\infty}\Sigma_N^{\dagger-1}(\boldsymbol{\vartheta})\Omega_N^{\dagger}(\boldsymbol{\vartheta})\Sigma_N^{\dagger-1}(\boldsymbol{\vartheta})\big], \tag{2.25}$$

where $\Sigma_N^{\dagger}(\boldsymbol{\vartheta}) = -\frac{1}{N}\mathrm{E}[\frac{\partial}{\partial\boldsymbol{\theta}^{\triangle\prime}}S_N^{\dagger}(\boldsymbol{\vartheta})]$ and $\Omega_N^{\dagger}(\boldsymbol{\vartheta}) = \frac{1}{N}\mathrm{E}[S_N^{\dagger}(\boldsymbol{\vartheta})S_N^{\dagger\prime}(\boldsymbol{\vartheta})]$, with the limit of the former being a positive definite matrix, and that of the latter simply a constant matrix. The matrix $\Sigma_N^{\dagger}(\boldsymbol{\vartheta})$ can be consistently estimated by $\frac{1}{N}H_N^{\dagger}(\hat{\boldsymbol{\vartheta}}_{\mathtt{AQS}})$, where $H_N^\dagger(\boldsymbol{\vartheta}) = -\frac{\partial}{\partial \boldsymbol{\theta}^{\triangle\prime}} S_N^\dagger(\boldsymbol{\vartheta})$ is the negative Hessian matrix. Denoting $\dot{\bar{F}}^{\circ}_{jN,\rho_k}(\delta) = \frac{\partial}{\partial \rho_k}\bar{F}^{\circ}_{jN}(\delta) = \bar{F}_{jN}G_k - G_k\bar{F}_{jN} - \mathtt{diag}(\bar{F}_{jN}G_k - G_k\bar{F}_{jN})$, by the same notational covensions the elements of $H_N^\dagger(\boldsymbol{\vartheta})$ are:

$$
\begin{aligned}
H^\dagger_{\beta\boldsymbol{\vartheta}}(\boldsymbol{\vartheta}) &= \big[\mathbb{X}'\mathbb{X},\ \{\mathbb{X}'\bar{F}_{jN}\mathbb{Y}\},\ \{\mathbb{X}'G^s_{kN}\varepsilon_N\}\big],\\
H^\dagger_{\lambda\beta}(\boldsymbol{\vartheta}) &= \{\mathbb{X}'\bar{F}^{\circ}_{jN}\mathbb{Y}\}',\\
H^\dagger_{\lambda\lambda}(\boldsymbol{\vartheta}) &= \{\mathbb{Y}_N'\bar{F}'_{j'N}\bar{F}^{\circ}_{jN}\mathbb{Y} + \varepsilon_N'[\mathtt{diag}(\bar{F}_{jN}\bar{F}_{j'N}) - \mathtt{diag}(\bar{F}_{jN})\bar{F}_{j'N}]\mathbb{Y}\},\\
H^\dagger_{\lambda\rho}(\boldsymbol{\vartheta}) &= \{\varepsilon_N'[G'_{kN}\bar{F}^{\circ}_{jN} - \dot{\bar{F}}^{\circ}_{jN,\rho_k} + \bar{F}^{\circ}_{jN}G_{kN}]\mathbb{Y}\},\\
H^\dagger_{\rho\beta}(\boldsymbol{\vartheta}) &= \{\mathbb{X}'(G^{\circ}_{kN} + G^{\circ\prime}_{kN})\varepsilon_N\}',\\
H^\dagger_{\rho\lambda}(\boldsymbol{\vartheta}) &= \{\varepsilon_N'(G^{\circ}_{kN} + G^{\circ\prime}_{kN})\bar{F}_{jN}\mathbb{Y}\}',\\
H^\dagger_{\rho\rho}(\boldsymbol{\vartheta}) &= \{\varepsilon_N'[G'_{kN}G_{k'N} - 2\mathtt{diag}(G_{kN}) + \mathtt{diag}(G_{kN}G_{k'N})]\varepsilon_N\}.
\end{aligned}
$$

Using the special case of Lemma B.3 where the diagonal elements of the matrices involved are zero, we obtain the VC matrix $N\Omega_N(\boldsymbol{\vartheta})$, having the distinct elements:

$$
\begin{aligned}
N\Omega^\dagger_{\beta\boldsymbol{\vartheta}}(\boldsymbol{\vartheta}) &= \big[\mathbb{X}'H_N\mathbb{X},\ \{\mathbb{X}'H_N\bar{F}^{\circ}_{jN}\mathbb{X}\boldsymbol{\beta}\},\ \{0\}\big],\\
N\Omega^\dagger_{\lambda\lambda}(\boldsymbol{\vartheta}) &= \{\mathrm{tr}\big(H_N\bar{F}^{\circ}_{jN}(H_N\bar{F}^{\circ}_{j'N} + H_N\bar{F}^{\circ\prime}_{j'N})\big) + \boldsymbol{\beta}'\mathbb{X}'\bar{F}^{\circ\prime}_{jN}H_N\bar{F}^{\circ}_{j'N}\mathbb{X}\boldsymbol{\beta}\},\\
N\Omega^\dagger_{\lambda\rho}(\boldsymbol{\vartheta}) &= \{\mathrm{tr}\big(H_N\bar{F}^{\circ}_{jN}(H_NG^{\circ}_{kN} + H_NG^{\circ\prime}_{kN})\big)\},\\
N\Omega^\dagger_{\rho\rho}(\boldsymbol{\vartheta}) &= \{\mathrm{tr}\big(H_NG^{\circ}_{kN}(H_NG^{\circ}_{k'N} + H_NG^{\circ\prime}_{k'N})\big)\}.
\end{aligned}
$$

The fact that the diagonal elements of the matrices in linear-quadratic forms makes the terms involving skewness and excess kurtosis vanish, left with only the heteroskedasticity matrix H_N. Therefore, it is still possible to obtain a consistent estimator of $\Omega_N(\boldsymbol{\vartheta})$ by plug-in method. Specifically, $\mathtt{diagv}(H_N)$ can be estimated by $\hat{\varepsilon}_{\mathrm{R}} \odot \hat{\varepsilon}_{\mathrm{R}}$, where $\hat{\varepsilon}_{\mathrm{R}}$ is the plug-in estimator of ε_N and $\odot$ represents the Hadamard product.

2.5.2 Case of large C or T

Consider now the cases where C or T or both grow with N. For the CQS function $S^c_N(\theta)$ given in (2.13), it is easy to see that its β and σ^2 components have zero expectations whether the errors are homoskedastic or heteroskedastic, and therefore they do not need to be further adjusted. As in Liu and Yang (2015, 2020), we make adjustments on the stochastic terms of the λ and ρ components, so that the adjusted functions become unbiased under unknown heteroskedasticity. We exclude the part related to σ^2 in inference for the same reason as in section 2.5.1, and denote the remaining parameters as $\vartheta = (\beta', \lambda', \rho')'$.

For the λ elements of $S_N^c(\theta)$, note that $\tilde{\varepsilon}_N'(\beta,\delta)B_N(\rho)W_{\ell j}Y_N = \varepsilon_N'\mathbb{Q}_2\bar{F}_{jN}\mathbb{Y}$ at the true parameter values. We have $\mathrm{E}(\varepsilon_N'\mathbb{Q}_2\bar{F}_{jN}\mathbb{Y}) = \mathrm{tr}(H_N\mathbb{Q}_2\bar{F}_{jN}) = \mathrm{tr}[H_N\mathrm{diag}(\mathbb{Q}_2\bar{F}_{jN})] = \mathrm{tr}[H_N\mathbb{Q}_2 \times$
$\mathtt{diag}(\mathbb{Q}_2)^{-1}\mathtt{diag}(\mathbb{Q}_2\bar{F}_{jN})] = \mathrm{E}(\varepsilon_N'\mathbb{Q}_2\bar{F}_{jN}^d\mathbb{Y})$, where $\bar{F}_{jN}^d = \mathtt{diag}(\mathbb{Q}_2)^{-1}\mathtt{diag}(\mathbb{Q}_2\bar{F}_{jN})$. Taking the difference between the terms inside the first and the last expectations, we obtain AQS functions for the λ elements robust against H_N:

$$\tilde{\varepsilon}_N'(\beta,\delta)[\bar{F}_{jN}(\delta) - \bar{F}_{jN}^d(\delta)]\mathbb{Y}(\delta), \;\; j = 1,\ldots,p.$$

For the ρ elements of $S_N^c(\theta)$, note that at the true parameters, $\tilde{\varepsilon}_N'(\beta,\delta)G_{kN}(\rho)\tilde{\varepsilon}_N(\beta,\delta) = \varepsilon_N'\mathbb{Q}_2G_{kN}\mathbb{Q}_2\varepsilon_N = \varepsilon_N'\mathbb{Q}_2G_{kN}\mathbb{Q}_2(\mathbb{Y}-\mathbb{X}_1\beta)$. We have $\mathrm{E}(\varepsilon_N'\mathbb{Q}_2G_{kN}\mathbb{Q}_2\varepsilon_N) = \mathrm{tr}(H_N\mathbb{Q}_2G_{kN}\mathbb{Q}_2) = \mathrm{tr}[H_N\mathtt{diag}(\mathbb{Q}_2G_{kN}\mathbb{Q}_2)] = \mathrm{E}(\varepsilon_N'\mathbb{Q}_2\mathtt{diag}(\mathbb{Q}_2)^{-1}\mathtt{diag}(\mathbb{Q}_2G_{kN}\mathbb{Q}_2)(\mathbb{Y}-\mathbb{X}_1\beta))$. Similarly, we obtain AQS functions for the ρ elements robust against H_N:

$$\tilde{\varepsilon}_N'(\beta,\delta)[\bar{G}_{kN}(\rho) - \bar{G}_{kN}^d(\rho)](\mathbb{Y}(\delta) - \mathbb{X}_1(\rho)\beta), \;\; k = 1,\ldots,q,$$

where $\bar{G}_{kN} = G_{kN}\mathbb{Q}_2$ and $\bar{G}_{kN}^d = \mathtt{diag}(\mathbb{Q}_2)^{-1}\mathtt{diag}(\mathbb{Q}_2\bar{G}_{kN})$.

These together give the robust AQS function for $\vartheta = (\beta', \lambda', \rho')'$:

$$S_N^{\diamond}(\vartheta) = \begin{cases} \mathbb{X}_1'(\rho)\tilde{\varepsilon}_N(\beta,\delta), \\ \tilde{\varepsilon}_N'(\beta,\delta)\bar{F}_{jN}^{\diamond}(\delta)\mathbb{Y}(\delta), \quad\;\; j = 1,\ldots,p, \\ \tilde{\varepsilon}_N'(\beta,\delta)\bar{G}_{kN}^{\diamond}(\rho)(\mathbb{Y}(\delta) - \mathbb{X}_1(\rho)\beta), \;\; k = 1,\ldots,q, \end{cases} \tag{2.26}$$

where $\bar{F}_{jN}^{\diamond}(\delta) = \bar{F}_{jN}(\delta) - \bar{F}_{jN}^d(\delta)$, and $\bar{G}_{kN}^{\diamond}(\rho) = \bar{G}_{kN}(\rho) - \bar{G}_{kN}^d(\rho)$. Solving $S_N^{\diamond}(\vartheta) = 0$ gives the robust AQS estimator $\hat{\vartheta}_{\mathrm{R}}$ of ϑ. This can again be done by first solving analytically for β:

$$\hat{\beta}_{\mathrm{R}}(\delta) = [\mathbb{X}_1'(\rho)\mathbb{Q}_2(\rho)\mathbb{X}_1(\rho)]^{-1}[\mathbb{X}_1'(\rho)\mathbb{Q}_2(\rho)\mathbb{Y}(\delta)], \tag{2.27}$$

and then solving for λ and ρ numerically in the concentrated equations.

It can be shown under mild regularity conditions (in Appendix A) that $\hat{\theta}_{\mathrm{R}} -$

$\theta \xrightarrow{p} 0$ and

$$\sqrt{N}(\hat{\vartheta}_{\mathtt{R}} - \vartheta) \xrightarrow{D} N\big[0,\ \lim_{N\to\infty} \Sigma_N^{\diamond -1}(\vartheta)\Omega_N^{\diamond}(\vartheta)\Sigma_N^{\diamond -1}(\vartheta)\big], \tag{2.28}$$

where $\Sigma_N^{\diamond}(\vartheta) = -\frac{1}{N}\mathrm{E}[\frac{\partial}{\partial \theta^{\triangle\prime}} S_N^{\diamond}(\vartheta)]$ and $\Omega_N^{\diamond}(\vartheta) = \frac{1}{N}\mathrm{E}[S_N^{\diamond}(\vartheta)S_N^{\diamond\prime}(\vartheta)]$, with the limit of the former being a positive definite matrix, and that of the latter simply a constant matrix.

Again, $\Sigma_N^{\diamond}(\vartheta)$ can be estimated by its sample counterpart, $\frac{1}{N}H_N^{\diamond}(\hat{\vartheta}_{\mathtt{AQSH}})$, where $H_N^{\diamond}(\vartheta) = -\frac{\partial}{\partial \theta^{\triangle\prime}} S_N^{\diamond}(\vartheta)$ is the negative Hessian. Note $\frac{\partial}{\partial \rho_k}\mathtt{diag}(\mathbb{Q}_2)^{-1} = -\mathtt{diag}(\mathbb{Q}_2)^{-2} \times \mathtt{diag}(\frac{\partial}{\partial \rho_k}\mathbb{Q}_2(\rho))$, where $\frac{\partial}{\partial \rho_k}\mathbb{Q}_2(\rho) = \mathbb{Q}_2(\rho)G_{kN}(\rho)\mathbb{P}_2(\rho) + \mathbb{P}_2(\rho)G_{kN}'(\rho)\mathbb{Q}_2(\rho)$. Denote $\frac{\partial}{\partial \rho_k}\mathbb{Q}_2(\rho) = \mathbb{Q}_{2,\rho_k}(\rho)$, and $A^d = \mathtt{diag}(\mathbb{Q}_2)^{-1}\mathtt{diag}(\mathbb{Q}_2 A)$ for a square matrix A. The elements of $H_N^{\diamond}(\vartheta)$ are:

$$\begin{aligned}
H_{\beta\vartheta}^{\diamond}(\vartheta) &= \big[\mathbb{X}_1'\mathbb{Q}_2\mathbb{X}_1,\ \big\{\mathbb{X}_1'\mathbb{Q}_2\bar{F}_{jN}\mathbb{Y}\big\},\ \big\{\mathbb{X}_1'\mathbb{Q}_2 G_{kN}^s \tilde{\varepsilon}_N\big\}\big],\\
H_{\lambda\beta}^{\diamond}(\vartheta) &= \big\{\mathbb{X}_1'\mathbb{Q}_2\bar{F}_{jN}^{\diamond}\mathbb{Y}\big\}',\\
H_{\lambda\lambda}^{\diamond}(\vartheta) &= \big\{\mathbb{Y}'[\bar{F}_{j'N}'\mathbb{Q}_2\bar{F}_{jN}^{\diamond}]\mathbb{Y} + \tilde{\varepsilon}_N'[(\bar{F}_{jN}\bar{F}_{j'N})^d - \bar{F}_{jN}^d\bar{F}_{j'N}]\mathbb{Y}\big\},\\
H_{\lambda\rho}^{\diamond}(\vartheta) &= \big\{\tilde{\varepsilon}_N'[G_{kN}^s\mathbb{Q}_2\bar{F}_{jN}^{\diamond} - \bar{F}_{jN}^d G_{kN} + G_{kN}\bar{F}_{jN}^d - 2(G_{kN}\mathbb{P}_2)^d\bar{F}_{jN}^d\\
&\qquad + \mathtt{diag}(\mathbb{Q}_2)^{-1}\mathtt{diag}(G_{kN}'\mathbb{Q}_2\bar{F}_{jN} + \mathbb{Q}_2\bar{F}_{jN}G_{kN} - \mathbb{Q}_2 G_{kN}^s\mathbb{Q}_2\bar{F}_{jN})]\mathbb{Y}\big\},\\
H_{\rho\beta}^{\diamond}(\vartheta) &= \big\{\mathbb{X}_1'\mathbb{Q}_2\bar{G}_{kN}^{\diamond}(\mathbb{Y} - \mathbb{X}_1\beta) + \tilde{\varepsilon}_N'\bar{G}_{kN}^{\diamond}\mathbb{X}_1\big\}',\\
H_{\rho\lambda}^{\diamond}(\vartheta) &= \big\{\mathbb{Y}'[\bar{F}_{jN}'\mathbb{Q}_2\bar{G}_{kN}^{\diamond}](\mathbb{Y} - \mathbb{X}_1\beta) + \tilde{\varepsilon}_N'[\bar{G}_{kN}^{\diamond}\bar{F}_{jN}]\mathbb{Y}\big\},\\
H_{\rho\rho}^{\diamond}(\vartheta) &= \big\{\tilde{\varepsilon}_N'[(G_{k'N}^s\mathbb{Q}_2 - G_{k'N})\bar{G}_{kN}^{\diamond} - G_{kN}(G_{k'N}\mathbb{Q}_2 + \mathbb{Q}_{2,\rho_{k'}}) - 2(G_{k'N}\mathbb{P}_2)^d\bar{G}_{kN}^d\\
&\qquad + (G_{kN}^s\mathbb{Q}_{2,\rho_{k'}} + G_{kN}G_{k'N}\mathbb{Q}_2)^d + \bar{G}_{kN}^{\diamond}G_{k'N}](\mathbb{Y} - \mathbb{X}_1\beta)\big\}.
\end{aligned}$$

Again, using the special case of Lemma B.3, we obtain the VC matrix $\Omega_N^{\diamond}(\vartheta)$, having the following distinct elements,

$$
\begin{aligned}
N\Omega^{\diamond}_{\beta\vartheta}(\vartheta) &= \big[\mathbb{X}_1'\mathbb{Q}_2 H_N \mathbb{Q}_2 \mathbb{X}_1,\ \big\{\mathbb{X}_1'\mathbb{Q}_2 H_N \mathbb{Q}_2 \bar{F}^{\diamond}_{jN}\mathbb{X}\boldsymbol{\beta}\big\},\ \big\{\mathbb{X}_1'\mathbb{Q}_2 H_N \mathbb{Q}_2 \bar{G}^{\diamond}_{kN}\mathbb{X}_2\phi\big\}\big],\\
N\Omega^{\diamond}_{\lambda\lambda}(\vartheta) &= \big\{\mathrm{tr}\big(H_N\mathbb{Q}_2\bar{F}^{\diamond}_{jN}(H_N\mathbb{Q}_2\bar{F}^{\diamond}_{j'N}+H_N\bar{F}^{\diamond\prime}_{j'N}\mathbb{Q}_2)\big)\\
&\qquad + \boldsymbol{\beta}'\mathbb{X}'\bar{F}^{\diamond\prime}_{jN}\mathbb{Q}_2 H_N\mathbb{Q}_2\bar{F}^{\diamond}_{j'N}\mathbb{X}\boldsymbol{\beta}\big\},\\
N\Omega^{\diamond}_{\lambda\rho}(\vartheta) &= \big\{\mathrm{tr}\big(H_N\mathbb{Q}_2\bar{F}^{\diamond}_{jN}(H_N\mathbb{Q}_2\bar{G}^{\diamond}_{kN}+H_N\bar{G}^{\diamond\prime}_{kN}\mathbb{Q}_2)\big)\\
&\qquad + \boldsymbol{\beta}'\mathbb{X}'\bar{F}^{\diamond\prime}_{jN}\mathbb{Q}_2 H_N\mathbb{Q}_2\bar{G}^{\diamond}_{kN}\mathbb{X}_2\phi\big\},\\
N\Omega^{\diamond}_{\rho\rho}(\vartheta) &= \big\{\mathrm{tr}\big(H_N\mathbb{Q}_2\bar{G}^{\diamond}_{kN}(H_N\mathbb{Q}_2\bar{G}^{\diamond}_{k'N}+H_N\bar{G}^{\diamond\prime}_{k'N}\mathbb{Q}_2)\big)\\
&\qquad + \phi'\mathbb{X}_2'\bar{G}^{\diamond\prime}_{kN}\mathbb{Q}_2 H_N\mathbb{Q}_2\bar{G}^{\diamond}_{k'N}\mathbb{X}_2\phi\big\}.
\end{aligned}
$$

Similar to the case of fixed C and T, the VC matrix involves again only H_N. To find a consistent estimator for H_N, note that similar to Section 2.4, $\tilde{\varepsilon}_N = \mathbb{Q}_2(\mathbb{Y} - \mathbb{X}\boldsymbol{\beta}) = \mathbb{Q}_2\varepsilon_N$ can be consistently estimated in general by its RAQS counterpart that $\hat{\varepsilon}_{\mathtt{R}} = \mathbb{Q}_2(\hat{\rho}_{\mathtt{R}})[\mathbb{Y}(\hat{\delta}_{\mathtt{R}}) - \mathbb{X}_1(\hat{\rho}_{\mathtt{R}})\hat{\beta}_{\mathtt{R}}]$. Since $\mathrm{E}(\tilde{\varepsilon}_N \odot \tilde{\varepsilon}_N) = (\mathbb{Q}_2 \odot \mathbb{Q}_2)\mathtt{diagv}(H_N)$, an estimator for $\mathtt{diagv}(H_N)$ is given by $[\mathbb{Q}_2(\hat{\rho}_{\mathtt{R}}) \odot \mathbb{Q}_2(\hat{\rho}_{\mathtt{R}})]^{-}(\hat{\varepsilon}_{\mathtt{R}} \odot \hat{\varepsilon}_{\mathtt{R}})$, where $[\cdot]^{-}$ is the generalized inverse.

Finally, to consistently estimate $\Omega^{\diamond}_N(\vartheta)$, we also need to do a bias-correction similar to the large C or T case of the AQS estimation under homoskedastic errors. Terms that are quadratic in ϕ need to be bias-corrected, which are $\phi'\mathbb{X}_2'\bar{F}^{\diamond\prime}_{jN}\mathbb{Q}_2 H_N\mathbb{Q}_2\bar{F}^{\diamond}_{j'N}\mathbb{X}_2\phi$, $\phi'\mathbb{X}_2'\bar{F}^{\diamond\prime}_{jN}\mathbb{Q}_2 H_N\mathbb{Q}_2\bar{G}^{\diamond}_{kN}\mathbb{X}_2\phi$, and $\phi'\mathbb{X}_2'\bar{G}^{\diamond\prime}_{kN}\mathbb{Q}_2 H_N\mathbb{Q}_2\bar{G}^{\diamond}_{k'N}\mathbb{X}_2\phi$. These terms are embedded in the last elements of $N\Omega^{\diamond}_{\lambda\lambda}(\vartheta)$, $N\Omega^{\diamond}_{\lambda\rho}(\vartheta)$, and $N\Omega^{\diamond}_{\rho\rho}(\vartheta)$ separately. The plug-in estimator of the first term is seen to be

$$
\begin{aligned}
&\hat{\phi}'\mathbb{X}_2'\bar{F}^{\diamond\prime}_{jN}\mathbb{Q}_2 H_N\mathbb{Q}_2\bar{F}^{\diamond}_{j'N}\mathbb{X}_2\hat{\phi}\\
&= \phi'\mathbb{X}_2'\bar{F}^{\diamond\prime}_{jN}\mathbb{Q}_2 H_N\mathbb{Q}_2\bar{F}^{\diamond}_{j'N}\mathbb{X}_2\phi + \varepsilon_N'\mathbb{P}_2\bar{F}^{\diamond\prime}_{jN}\mathbb{Q}_2 H_N\mathbb{Q}_2\bar{F}^{\diamond}_{j'N}\mathbb{P}_2\varepsilon_N + o_p(N)\\
&= \phi'\mathbb{X}_2'\bar{F}^{\diamond\prime}_{jN}\mathbb{Q}_2 H_N\mathbb{Q}_2\bar{F}^{\diamond}_{j'N}\mathbb{X}_2\phi + \mathrm{tr}(H_N\mathbb{P}_2\bar{F}^{\diamond\prime}_{jN}\mathbb{Q}_2 H_N\mathbb{Q}_2\bar{F}^{\diamond}_{j'N}\mathbb{P}_2) + o_p(N). \qquad (2.29)
\end{aligned}
$$

Therefore, the plug-in estimator of the term $\boldsymbol{\beta}'\mathbb{X}'\bar{F}^{\diamond\prime}_{jN}\mathbb{Q}_2 H_N\mathbb{Q}_2\bar{F}^{\diamond}_{j'N}\mathbb{X}\boldsymbol{\beta}$ in $N\Omega^{\diamond}_{\lambda\lambda}(\vartheta)$ should be substracted for $\mathrm{tr}(H_N\mathbb{P}_2\bar{F}^{\diamond\prime}_{jN}\mathbb{Q}_2 H_N\mathbb{Q}_2\bar{F}^{\diamond}_{j'N}\mathbb{P}_2)$. Similarly, we need to subtract

$\mathrm{tr}(H_N\mathbb{P}_2\bar{F}^{\diamond\prime}_{jN}\mathbb{Q}_2 H_N\mathbb{Q}_2\bar{G}^{\diamond}_{kN}\mathbb{P}_2)$, and $\mathrm{tr}(H_N\mathbb{P}_2\bar{G}^{\diamond\prime}_{kN}\mathbb{Q}_2 H_N\mathbb{Q}_2\bar{G}^{\diamond}_{k'N}\mathbb{P}_2)$ for the plug-in estimators of the last elements in $N\Omega^{\diamond}_{\lambda\rho}(\vartheta)$, and $N\Omega^{\diamond}_{\rho\rho}(\vartheta)$.

2.6 Construction of Space-Time Connectivity Matrices

A crucial step in modelling the spatiotemporal data may be the construction of the space-time connectivity/weight matrices. There are interesting proposals and discussions in the literature (Pace et al. 1998a,b, 2000; Sun et al. 2005), but it is obvious that more rigorous ways for constructing these matrices are desired. As pointed out in the introduction, the real estate prices data are typically subject to dependence over space and time, cluster specific effects, and economic shocks over time. The space-time dependence typically decays with distance and time, a cluster may be a building or a neighborhood, and a economic shock may be a policy change or a financial crisis.

We view that non-negligible dependence between two spatial units exists only when they are 'nearby' in both space and time. Let δ be the distance threshold within which the spatial dependence is considered non-negligible, and τ_p and τ_c are the 'past' and 'current' time limits within which the temporal dependence is considered non-negligible.[9] Recall that $t_i, i = 1, \ldots N$, the times at which the prices are collected. Let d_{ij} be the distance between units i and j. Define a space-dominated space-time connectivity matrix W_d with elements:

$$w_{ij} = f_d(d_{ij})\mathbf{1}\{d_{ij} \leq \delta\}\mathbf{1}\{|t_i - t_j| \leq \tau_c \text{ or } \tau_c \leq t_i - t_j \leq \tau_p\}, \tag{2.30}$$

where $\mathbf{1}\{\cdot\}$ is the indicator function. Similarly, we can also define a time-dominated space-time connectivity matrix W_t with elements:

$$w_{ij} = f_t(|t_i - t_j|)\mathbf{1}\{d_{ij} \leq \delta\}\mathbf{1}\{|t_i - t_j| \leq \tau_c \text{ or } \tau_c \leq t_i - t_j \leq \tau_p\}, \tag{2.31}$$

where $f_d(\cdot)$ and $f_t(\cdot)$ are functions capturing the attenuation patterns.[10]

In our empirical application, we consider two different space-dominated space-time connectivity matrices that help capture the spatiotemporal interaction effects between housing transactions. The first connectivity matrix $W_{d,cmty}$ aims to capture the interdependence within a community. Figure 2.4 shows the transactions by

[9] Another reason to assume a fixed space and time window is to make the space-time connectivity matrix a 'band' matrix, so that the computation burden is alleviated, as mentioned in Section 2.2.

[10] Different functional forms can be used such as exponential distance weights, power distance weights and inverse distance weights.

communities in the dataset. As the location information of the housing transactions are up to the community level, it is natural that we construct $W_{d,cmty}$ using transactions that are located in the same community. Suppose the $w_{cmty,ij}$ is an elements in $W_{d,cmty}$. A non-zero $w_{cmty,ij}$ indicates that there exists interdependence between unit i and j. We set two criteria for $w_{cmty,ij}$ to be non-zero as in (2.30).[11] First, transaction i and j are within the same community. This captures the interaction along the spatial dimension. Second, transaction j takes place before or after transaction i within one month time. This captures the interaction along the the temporal dimension. Thus, we define

$$w_{cmty,ij} = f_{d,cmty}(d_{ij})\mathbf{1}\{Cmty(i) = Cmty(j)\}\mathbf{1}\{-30 \leq t_i - t_j \leq 30\},$$

where $f_{d,cmty}(d_{ij})$ is set to be 1 as the location information is only up to the community level. $Cmty(i)$ gives the community of transaction i, and t_i represents the calendar day in 2015 that transaction i took place.

The second connectivity matrix $W_{d,nbr}$ is designed to capture the interdependence between a housing unit and its neighbours outside of communities, e.g., 5 nearest neighbours (5NN) that are within 5 kilometers as our applications.[12] Thus $w_{nbr,ij}$ is defined as

$$w_{nbr,ij} = f_{d,nbr}(d_{ij})\mathbf{1}\{5\, Nearest\, Neighbours\, of\, i\, within\, 5\, km\}\mathbf{1}\{-30 \leq t_i - t_j \leq 30\},$$

where $f_{d,nbr}(d_{ij})$ is set to be a function of inverse distance $f_{d,nbr}(d_{ij}) = 1/distance(i,j)$. Figure 2.5 shows the nearest 5 neighbors outside of communities.

2.7 Monte Carlo Study

Monte Carlo experiments are conducted to investigate the finite sample performance of the QML, AQS, and RAQS estimators of the parameters in the $\mathtt{STAR}(p,q)$ model, with different sample sizes, different ways of including cluster and time effects, different error distributions, and homoskedasticity or heteroskedasticity.

[11] In the application, the space-time connectivity matrix is further row normalized. A non-zero element $w_{cmty,ij}$ equals to $1/n_i$, where n_i is the number of non-zero elements in row i, which represents the number of transactions that have effects on transaction i.

[12] We also considered the cases of 10NN and 15NN. The results are presented in Figure 2.8 and Figure 2.9.

For the data generating process (DGP), we adopt a STAR model (as in (2.4)) of order 2 for both spatial lag and spatial error:

$$A_N(\lambda)Y_N = \mathbf{X}_N\boldsymbol{\beta} + V_N, \;\; B_N(\rho)V_N = \varepsilon_N,$$

where $A_N(\lambda) = I_N - \sum_{j=1}^{2} \lambda_j W_{\ell j}$ and $B_N(\rho) = I_N - \sum_{k=1}^{2} \rho_k W_{ek}$.

We set the coefficients β to be $(1,2,3)'$, the spatial lag parameters $\lambda' = (0.5, 0.3)$, and the spatial error parameters $\rho' = (0.2, 0.4)$. We assume $\varepsilon_i \overset{i.i.d}{\sim} (0,1)$ for models of homoskedastic errors, and $\varepsilon_i \overset{i.n.i.d}{\sim} (0, h_{N,i})$ for models of heteroskedatic errors. Sample size N takes values from $\{400, 800, 1200\}$. $\mathbf{X}_N = [X_N, M_C, M_{T-1}]$. In the simulation, X_N are fixed regressors drawn from standard normal distribution. The number of clusters/time segments are assumed to be fixed or to increase with sample size. Each set of Monte Carlo results is based on 2,000 Monte Carlo samples. The processes of generating space-time connectivity matrices, and the way we assign clusters and time segments to observations are described below.

Space-Time Connectivity Matrices: Under homoskedasticity, we first construct the space-time connectivity matrices based on the queen and rook contiguity, where queen contiguity is followed for constructing W_{l1} and W_{e1}, and rook contiguity for constructing W_{l2} and W_{e2}. Under heteroskedasticity, we use a group interaction scheme as in Lin and Lee (2010) for constructing W_{lj} and queen contiguity for constructing W_{ek}.

Cluster-membership Matrix for Model with Additive Fixed Effects: We consider two scenarios for number of clusters/time segments included in the model. One is when the number of clusters and time segments are relatively small and do not grow with sample size (fixed), while the other is when the number of clusters and time segments are large and grow with sample size. Assume the probability that a individual transactions belongs to a cluster follows a multinomial distribution. If we further assume each observation belongs to a cluster with equal probability, and we assign each observation to C potential clusters, we will have N/C observations in each cluster. We also assume the the time segments are equally sized. For example, suppose there are 3 time segments, the first third of the sample belongs to the first time segment, the second third of the sample belongs to the second time segment, and the last third of the sample belongs to the third segment.

Below is an example showing what matrix M_C and M_T look like when sample size is 6 and there are 3 clusters and 2 time segments.

$$M_C = \begin{bmatrix} \mathbf{0} & \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} \\ \mathbf{1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} & \mathbf{0} \\ \mathbf{1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} \end{bmatrix}, \ M_T = \begin{bmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{1} & \mathbf{0} \\ \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} \\ \mathbf{0} & \mathbf{1} \\ \mathbf{0} & \mathbf{1} \end{bmatrix}$$

Error Distribution: For the estimation of the QMLE and AQSE under homoskedasticity, we consider three error distributions: (i) ε follows an $i.i.d.$ $N(0,1)$ distribution, (ii) ε follows a mixture normal distribution, and (iii) ε follows an chi-squared distribution. The normal mixture still gives an symmetric distribution, but it is leptokurtic. The chi-squared distribution is both skewed and leptokurtic.[13] Since we know that QML/AQS estimators are robust against non-normality, we expect the simulation results to be consistent in all three scenarios when clusters and time segments are small/large. For the estimation of the robust AQSE under heteroskedasticity, the idiosyncratic errors are generated in a similar fashion as in Lee and Yu (2010b).[14]. The simulation results are reported from Table 2.1 to Table 2.3.B.

Simulation Results: Table 2.1 presents partial Monte Carlo results for the QML estimation of the $\texttt{STAR}(2,2)$ model, where the number of clusters and time segments are fixed and the errors are homoskedastic. The results show excellent performance of the QML estimators of the model parameters. The consistency of the QML estimators are clearly demonstrated.

Table 2.2 presents partial Monte Carlo results the for the QML and AQS estimators of the $\texttt{STAR}(2,2)$ model, where the number of clusters and time segments

[13]The normal mixture random variates are generated through $\varepsilon_i = ((1-\xi_i)Z_i + \xi_i \tau Z_i) / \left(1-p+p*\tau^2\right)^{0.5}$, where $Z_i \sim N(0,1)$, and $\xi_i \sim$ Bernoulli(p) independent of Z_i. The standardized chi-squared variates are given by $\varepsilon_i = (Q_i-3)/6^{0.5}$, where Q_i follows a chi-squared distribution with degree of freedom 3. We choose $p = 0.1$, meaning that 90% of the random variates are from standard normal and the remaining 10% are from another normal population with standard deviation τ. We set $\tau = 4$ in our Monte Carlo experiments.

[14]If the group size is larger than the average group size, the variance is set to be the same as the group size, otherwise, the variance is the square of the inverse of the group size

increase with sample size and the errors are homoskedastic. When the number of clusters and time segments increase with sample, the regular QML estimators are no longer consistent. This is clearly demonstrated by the results in Table 2.2, in particular the QMLEs of the spatial parameters as shown in Table 2.2.A and the QMLE of the error variance as shown in Table 2.2.B. In contrast, the AQS estimator addresses the incidental parameters problem by adjusting the concentrated quasi score functions, and therefore they are consistent and its excellent finite sample performance is clearly demonstrated in Table 2.2.

Table 2.3 presents partial Monte Carlo results for the AQS and RAQS estimators of the $\mathtt{STAR}(2,2)$ model where the number of clusters and time segments increase with sample size and the errors are heteroskedastic. The results show an excellent finite sample performance of the RAQS estimators in terms of Monte Carlo bias, sd and the estimated se. In this case, the finite sample performance of the AQS estimators seems fine in terms of bias, but not in terms of standard error estimation. The standard error estimation under AQS framework seems to be worse for the spatial parameters as seen in Table 2.3.A than for the regression coefficients and error variance as seen in Table 2.3.B.

Addtional Monte Carlo results (unreported for brevity) show that the QMLE performs well only when C and T are fixed and the errors are homskedastic, AQSE performs well in general when errors are homoskedastic, and the RAQSE performs well in all situations considered whether C and T are small or large, and whether the errors are homoskedastic or heteroskedastic. In case of homoskedastic errors, the AQSE has an advantage of being able to conduct inference (estimation, confidence interval and test) for the error variance.

2.8 An Empirical Application

We provide an application using the resale housing transaction data in Beijing in 2015. The dataset has detailed information on housing transactions such as the exact location of the house/apartment at the community level, type of the building, age of the building, transaction date, etc. These detailed information allows us to define the clusters and time segments from different dimensions, thus enables us to have a good design of the space-time connectivity matrices to capture the spatiotemporal effects and to have flexibla ways of controlling cluster and time effects. These features greatly fit into our specification setup.

2.8.1 Data, Variables and Summary Statistics

We collect the resale housing transaction data from Lianjia.com.[15] Our dataset consists of 103149 micro-data of the housing transactions in 11 out of the 16 districts in Beijing. It covers 52.2% of all the resale housing transactions in Beijing in that year. Figure 2.1 shows the map of Beijing together with the districts covered by the data set.

We have information about price, location, and housing attributes such as living area, number of bedroom/kitchen/bathroom, decoration, and age. It also includes the building characteristics such as building type, the structure type, and whether a building has an elevator or not.[16] Table 2.4 provides the summary statistics for the variables in the data set.

2.8.2 Clusters and Time Segments

The existence of the heterogeneity across clusters and time segments are evident, which gives good justification for including cluster and time segment effects. Figure 2.2 shows the district level monthly average transaction housing prices in Beijing. We can see that districts closer to the city center have higher housing prices. Also, note that housing prices are different across time segments, and different districts show different trends. These evidence inspires us to include District $\times$ Month effects to control for the district-month heterogeneity. In this scenario,

[15]Lianjia Real Estate Agency Co. Ltd is the largest real estate agency in China. In China, most of the transactions are conducted via real estate agencies, where Lianjia takes up the largest market share. In 2018, Lianjia takes up 49% of the resale housing market in Beijing (China Galaxy Securities 2019)(http://pdf.dfcfw.com/pdf/H3_AP201907071337986075_1.pdf).

[16]Tower-type and Slab-type are the two major tall-building types for the apartments in China. For tower-type buildings, there are usually more than four or five housing units surrounding the elevators placed in the center of a building, and thus the units are usually not north-south transparent. Slab-type buildings are in shape of rectangles and usually north-south transparent. Typically apartments in slab-type buildings have higher prices. Structure type can be brick-concrete, steel-concrete or mixed.

11 districts and 12 months gives $11 \times 12 = 132$ clusters, which is relatively small compared to our sample size. Figure 2.3 shows heterogeneity of the housing prices across communities. Larger bubbles in Figure 2.3 are associated with higher housing prices. Consistent with what is shown in Figure 2.2, communities that are closer to the city center usually have higher housing prices, which makes it plausible to include community fixed effects. In the dataset, we have around 5,000 communities in 2015, and around 2,400 of them have more than 10 transactions.[17] The number of communities are relatively large compared to the sample size. In this case, we adopt the AQS and the RAQS estimators discussed in the section 2.4 and 2.5 to obtain consistent estimates of the spatiotemporal interaction effects, controlling for large community fixed effects and unknown heteroskedasticity.

2.8.3 Space-Time Connectivity Matrices

The general ways of constructing the space-time connectivity matrices have been discussed in Sec. 2.6 in connection to this data.

2.8.4 STAR Model Specification

We consider the $\mathtt{STAR}(1,0)$, $\mathtt{STAR}(2,0)$, and $\mathtt{STAR}(2,2)$ models in this section. We adopt different estimation methods to guarantee consistency, depending on the ways of including cluster and time segment effects.

First, we consider the specifications with district effects and week effects. It is similar to Equation (2.1), taking the following form:

$$A_N(\lambda)Y_N = X_N\beta + M_{dist}\mu + M_{week}\alpha + V_N, \;\; B_N(\rho)V_N = \varepsilon_N, \tag{2.32}$$

where for $\mathtt{STAR}(1,0)$, $B_N(\rho) = I_N$, and $A_N(\lambda) = I_N - \lambda_{cmty}W_{d,cmty}$ or $I_N - \lambda_{nbr}W_{d,nbr}$; for $\mathtt{STAR}(2,0)$, $B_N(\rho) = I_N$, and $A_N(\lambda) = I_N - \lambda_{cmty}W_{d,cmty} - \lambda_{nbr}W_{d,nbr}$; and for $\mathtt{STAR}(2,2)$, $A_N(\lambda) = B_N(\rho) = I_N - \lambda_{cmty}W_{d,cmty} - \lambda_{nbr}W_{d,nbr}$.

In this empirical study, Y_N is the log of total price. Following the literature of hedonic housing models, X_N are the housing attributes including living area, squared living area, number of bedrooms, living rooms, kitchens, bathrooms, and the deco-

[17] In the application, we consider community fixed effects for communities with more than 10 transactions.

ration, age, etc., as shown in Table 2.4. M_{dist} is the district membership matrix, and μ are the district fixed effects. M_{week} is the week membership matrix, and α are the week fixed effects. $W_{d,cmty}$ and $W_{d,nbr}$ are defined in Section 2.6. In the case of including district and week effects, the number of clusters are small relative to the sample size, thus QML method is able to provide us with the consistent estimates.

Second, we consider the specification with community and week fixed effects. The specification is given by

$$A_N(\lambda)Y_N = X_N\beta + M_{cmty}\mu + M_{week}\alpha + V_N, \;\; B_N(\rho)V_N = \varepsilon_N. \tag{2.33}$$

μ are the community fixed effects. Other terms are the same as in (2.32). Since the number of communities (around 2,400) are large relative to the sample size, we adopt the AQS estimators. We further adopt the RAQS estimators to address that ε_N can potentially be heteroskedastic. Unfortunately, using the whole sample (about 10,000 $\times$ 10,000) causes memory problems and the computation can be very slow. For this reason, we run regression after re-grouping the data every two months and produce the results separately.[18] After re-grouping the data, we are able to conduct all the analysis using a computer with 64 GB of RAM. Another advantage of doing this re-grouping is that it enables us to check the robustness of our estimation method, which we will discuss in section 2.8.5

2.8.5 Results

Under specification (2.32) and (2.33), we first obtain the estimation results for the spatialtemporal interaction effects within communities (λ_{cmty}) estimated from the $\texttt{STAR}(1,0)$ and $\texttt{STAR}(2,0)$ models, as shown in Figure 2.6.A. The estimation results of λ_{cmty} are all significant. However, there are huge difference between the two estimates. If we do not control for community fixed effects, λ_{cmty} will be severely underestimated. On average, the magnitude of the AQS estimates of λ_{cmty} are about 1.5 times larger than the QML estimates. Note that in Figure 2.6.A, we assume there is no heterskedasticity in the model. If we further allow for the

[18]We do not use the data in January and February, because the sample size in these two months are significantly smaller than the rest, and 10 days of the data in February are missing. There is no missing data problem from March onward. Table 2.10 reports the sample size in each month.

existence of the unknown heteroskedasticity, we have the RAQS estimates in Figure 2.6.B. There are still significant difference between the QML and RAQS estimates.

Figure 2.7.A and Figure 2.7.B are the estimation of the spatiotemporal interaction effects from 5 nearest neighbours outside of communities (λ_{nbr}). Figure 2.7.A shows that we will sometimes slightly overestimate λ_{nbr} if we do not control for the community fixed effects, e.g. with May-Jun and Jul-Aug samples.[19] Figure 2.7.B shows that the estimation of λ_{nbr} will be much (around 40%) lower if we do not allow for unknown heteroskedasticity.

Comparing Panel A and Panel B in Figure 2.6.A – Figure 2.7.B, we can see that the estimation results are very robust across the $\texttt{STAR}(1,0)$ and $\texttt{STAR}(2,0)$ specifications.

Table 2.5 reports the above mentioned figure plots of λ_{cmty} and λ_{nbr} of the $\texttt{STAR}(1,0)$, $\texttt{STAR}(2,0)$ models in a table, with additional results from $\texttt{STAR}(2,2)$ model. We also reports the coefficients for other housing attributes in the Table 2.6 – Table 2.9.

2.8.6 Computational Notes

Estimating the parameters and implementing the inference procedures can be very time-consuming when dealing with large data. The most time-consuming matrix operations are the inversion and multiplication. Here we share some experience on the ways to speed up the computations. We use Matlab 2021a for the empirical application, but the experience should also apply to other programming languages.

First, we need to utilize the sparsity of the space-time connectivity matrices, and try to avoid the calculation of the inverse of large dense matrices. We suggest to use the user-written function `pseudoinverse`[20] to calculate the generalized inverse matrix. It implements the Moore-Penrose pseudoinverse factorization on a matrix that is later used for matrix multiplications. The good features of this function are that it can deal with the pseuinverse of a sparse matrix, and it does not involve singular

[19]We find similar patterns when we define the connectivity matrix using 10 or 15 nearest neighbours. The estimation results are very robust, which are presented in Figure 2.8 and Figure 2.9.

[20]For the documentation of this function, please refer to https://www.mathworks.com/matlabcentral/f-ileexchange/25453-pseudo-inverse

value decomposition, which are different from the build-in matrix-pseudoinverse function `pinv`. This makes the calculation much faster. For example, the computation time of the pseudoinverse of A_N in the $\texttt{STAR}(2,0)$ model (for March-April sample with around 18,000 observations, see Table 2.10) using `pseudoinverse` function is around 2.3 seconds, while it takes more than 2,800 seconds if we use `pinv`.

Second, we should try to minimize the frequency of calculating dense matrix multiplications, which appear frequently in the model inference parts. The multiplication of large dense matrices are also very time-consuming. For this we offer three tips:

(i) Avoid repetitive computation. Some matrix multiplications in the elements of the Hessian matrix and VC matrix appear repeatedly. We should pre-define them, do the calculation and store the multiplications first, instead of doing repetitive calculation everytime they appear. Second,

(ii) Avoid the use of the build-in `trace` function if the goal is to obtain the trace of a matrix multiplication. This is because we only need the diagonal elements to calculate the trace. For example, to calculate the trace of the multiplication of two matrices A and B, one should use the Matlab command $\texttt{sum}(\texttt{sum}(A'.*B))$ instead of $\texttt{trace}(A*B)$, where the former takes around 12 seconds and the latter takes around 236 seconds when the sizes of A and B are around 18,000 $\times$ 18,000.

(iii) Use element-wise product of vectors instead of matrix multiplication when calculating the product of two diagonal matrices, as the results only depend on the product of the diagonal elements. This is relevant when implementing the inference procedures of the RAQS method.

Lastly, we note that storing the large matrices created in the intermediate steps takes up large memory space. To save memory space, first, do not store a product of matrices as a new matrix if it only appears once, and second, clean up the matrices that are no longer in use in the later steps. For the empirical applications considered in this paper, the largest sample size is around 24,000 (November-December sample), which requires close to 60GB of RAM. All the results can be produced using a workstation with 64GB of RAM.

2.9 Conclusion and Discussion

Estimating the interaction effects or spillover effects is an very important topic in the housing literature. Yet, despite the great features of spatiotemporal data, spatial econometric theories and methods seem lag behind for sophisticated analyses of spatiotemporal data. In this paper, we propose a high-order spatiotemporal autoregressive (STAR) model with unobserved cluster and time heterogeneity to study the spatiotemporal interaction effects in the housing market. We propose an *adjusted quasi score* (AQS) method that allows us to have consistent estimates and asymptotic normality for inference when the clusters and/or time segments grow with sample size. When there exists unknown heteroskedasticity, a *robust adjusted quasi score* (RAQS) method is proposed.

Our Monte Carlo results demonstrates an excellent finite sample performance of the QML, AQS and RAQS estimators when the number of clusters (C) and the number of time segements (T) are fixed and the errors are homoskedastic, an excellent finite sample performance of the AQS and RAQS estimators when C and/or T grow with sample size N the errors are homoskedastic, and an excellent finite sample perfor of the RAQS estimator when the errors are heteroskedastic. We can see the AQS and RAQS estimators are especially useful when we have large unobserved cluster and time heterogeneity, and when there exists heteroskedasticity in the error terms.

Our proposed estimation and inference methods are applied to the housing transaction data in Beijing to study the spatiotemporal interaction effects. We find that there exists significant positive interaction effects, through both within communities transactions and neighbours outside of communities. However, the estimation of the interaction effects are largely affected after controlling for cluster heterogeneity at the community level. If we do not allow for community heterogeneity, we are likely to underestimate the interaction effects within communities, and overestimate the interaction effects from neighbours that are outside of communities. If we further allow for heteroskedasticity in the errors, the difference between RAQSE and QMLE are especially large for the interaction effects from neighbours that are outside of communities. These finds provides good justifications for using AQS/RAQS methods for estimating the spatiotemproal interaction effects when there exists unobserved cluster and time heterogeneity and unknown heteroskedasticity.

While our empirical analysis illustrates the importance of allowing for cluster and time heterogeneity in the STAR model, future work would benefit from further explorations on the relationship between the bias and the unobserved cluster and time heterogeneity. Taking a closer look at this question will help us understand the reason why we would over/underestimate the spatiotemporal interaction effects.

References

Anselin, L. and Cho, W. K. T. (2002). Spatial Effects and Ecological Inference. *Political Analysis*, 10(3):276–297.

Anselin, L. and Lozano-Gracia, N. (2008). Errors in Variables and Spatial Effects in Hedonic House Price Models of Ambient Air Quality. *Empirical Economics*, 34(1):5–34.

Baltagi, B. H., Bresson, G., and Etienne, J.-M. (2015). Hedonic Housing Prices in Paris: an Unbalanced Spatial Lag Pseudo-Panel Model with Nested Random Effects. *Journal of Applied Econometrics*, 30(3):509–528.

Campbell, J. Y., Giglio, S., and Pathak, P. (2011). Forced Sales and House Prices. *American Economic Review*, 101(5):2108–31.

China Galaxy Securities (2019). 2019 Report on the Real Estate Agency Industry.

Dorsey, R. E., Hu, H., Mayer, W. J., and Wang, H.-c. (2010). Hedonic versus Repeat-Sales Housing Price Indexes for Measuring the Recent Boom-Bust Cycle. *Journal of Housing Economics*, 19(2):75–93.

Gerardi, K., Rosenblatt, E., Willen, P. S., and Yao, V. (2012). Foreclosure Externalities: Some New Evidence. *NBER Working Paper*.

Goodman, A. C. and Thibodeau, T. G. (1995). Age-related Heteroskedasticity in Hedonic House Price Equations. *Journal of Housing Research*, pages 25–42.

Goodman, A. C. and Thibodeau, T. G. (1997). Dwelling-age-related Heteroskedasticity in Hedonic House Price Equations: An Extension. *Journal of Housing Research*, pages 299–317.

Gupta, A. (2019). Foreclosure Contagion and the Neighborhood Spillover Effects of Mortgage Defaults. *The Journal of Finance*, 74(5):2249–2301.

Holly, S., Pesaran, M. H., and Yamagata, T. (2010). A Spatio-Temporal Model of House Prices in the USA. *Journal of Econometrics*, 158(1):160–173.

Holly, S., Pesaran, M. H., and Yamagata, T. (2011). The Spatial and Temporal Diffusion of House Prices in the UK. *Journal of Urban Economics*, 69(1):2–23.

Honkanen, P. and Schmidt, D. (2017). Price and Liquidity Spillovers during Fire Sale Episodes. *HEC Paris Research Paper No. FIN-2017-1214*.

Lancaster, T. (2000). The Incidental Parameter Problem since 1948. *Journal of Econometrics*, 95(2):391–413.

Lee, L.-f. and Yu, J. (2010a). Estimation of Spatial Autoregressive Panel Data Models with Fixed Effects. *Journal of Econometrics*, 154(2):165–185.

Lee, L.-f. and Yu, J. (2010b). Some Recent Developments in Spatial Panel Data Models. *Regional Science and Urban Economics*, 40(5):255–271.

Lin, X. and Lee, L.-f. (2010). GMM Estimation of Spatial Autoregressive Models with Unknown Heteroskedasticity. *Journal of Econometrics*, 157(1):34–52.

Liu, S. F. and Yang, Z. (2015). Modified QML Estimation of Spatial Autoregressive Models with Unknown Heteroskedasticity and Nonnormality. *Regional Science and Urban Economics*, 52:50–70.

Liu, S. F. and Yang, Z. (2017). Heteroskedasticity Robust Estimation and Testing for Higherorder Spatial Autoregressive Models. *Working Paper*.

Liu, S. F. and Yang, Z. (2020). Robust Estimation and Inference of Spatial Panel Data Models with Fixed Effects. *Japanese Journal of Statistics and Data Science*, 3(1):257–311.

Maclennan, D. (2012). Understanding Housing Markets: Real Progress or Stalled Agendas? In Clapham, D. F., Clark, W. A., and Gibb, K., editors, *The Sage Handbook of Housing Studies*. Sage.

Nappi-Choulet Pr, I. and Maury, T.-P. (2009). A Spatiotemporal Autoregressive Price Index for the Paris Office Property Market. *Real Estate Economics*, 37(2):305–340.

Neyman, J. and Scott, E. L. (1948). Consistent Estimates Based on Partially Consistent Observations. *Econometrica*, pages 1–32.

Pace, R. K., Barry, R., Clapp, J. M., and Rodriquez, M. (1998a). Spatiotemporal Autoregressive Models of Neighborhood Effects. *The Journal of Real Estate Finance and Economics*, 17(1):15–33.

Pace, R. K., Barry, R., Gilley, O. W., and Sirmans, C. (2000). A Method for Spatial-Temporal Forecasting with an Application to Real Estate Prices. *International Journal of Forecasting*, 16(2):229–246.

Pace, R. K., Barry, R., and Sirmans, C. F. (1998b). Spatial Statistics and Real Estate. *The Journal of Real Estate Finance and Economics*, 17(1):5–13.

Stevenson, S. (2004). New Empirical Evidence on Heteroscedasticity in Hedonic Housing Models. *Journal of Housing Economics*, 13(2):136–153.

Sun, H., Tu, Y., and Yu, S.-M. (2005). A Spatio-Temporal Autoregressive Model for Multi-unit Residential Market Analysis. *The Journal of Real Estate Finance and Economics*, 31(2):155–187.

Yang, Z. (2018). Bootstrap LM Tests for Higher Order Spatial Effects in Spatial Linear Regression Models. *Empirical Economics*, 55:35–68.

Table 2.1: Empirical Mean(sd)[$\hat{\text{se}}$] of QMLE: $\texttt{STAR}(2,2)$ Model, C = T = 10, Homoskedasticity

n	$\delta_0 = (\lambda_0, \rho_0)'$	Normal	Normal Mixture	Chi-Squared
		Spatiotemporal Effects		
400	$\lambda_1 = 0.5$	0.498 (0.018) [0.017]	0.498 (0.018) [0.017]	0.499 (0.018) [0.017]
	$\lambda_2 = 0.3$	0.298 (0.016) [0.015]	0.298 (0.015) [0.015]	0.299 (0.015) [0.015]
	$\rho_1 = 0.2$	0.188 (0.080) [0.074]	0.188 (0.081) [0.074]	0.186 (0.082) [0.075]
	$\rho_2 = 0.4$	0.409 (0.062) [0.056]	0.408 (0.063) [0.056]	0.407 (0.060) [0.056]
800	$\lambda_1 = 0.5$	0.499 (0.014) [0.013]	0.499 (0.014) [0.013]	0.499 (0.014) [0.013]
	$\lambda_2 = 0.3$	0.299 (0.012) [0.012]	0.299 (0.012) [0.012]	0.299 (0.012) [0.012]
	$\rho_1 = 0.2$	0.193 (0.055) [0.053]	0.195 (0.056) [0.053]	0.194 (0.054) [0.053]
	$\rho_2 = 0.4$	0.405 (0.042) [0.040]	0.403 (0.040) [0.040]	0.405 (0.041) [0.040]
1200	$\lambda_1 = 0.5$	0.499 (0.010) [0.010]	0.499 (0.010) [0.010]	0.499 (0.010) [0.010]
	$\lambda_2 = 0.3$	0.300 (0.009) [0.009]	0.300 (0.009) [0.009]	0.300 (0.009) [0.009]
	$\rho_1 = 0.2$	0.196 (0.045) [0.044]	0.194 (0.045) [0.044]	0.196 (0.046) [0.044]
	$\rho_2 = 0.4$	0.403 (0.033) [0.033]	0.402 (0.032) [0.033]	0.403 (0.034) [0.033]
		Covariate Coefficients and Error Variance		
400	$\beta_1 = 1$	1.000 (0.073) [0.070]	1.000 (0.072) [0.070]	1.000 (0.075) [0.070]
	$\beta_2 = 2$	2.000 (0.067) [0.065]	1.999 (0.068) [0.065]	2.001 (0.067) [0.065]
	$\beta_3 = 3$	2.998 (0.075) [0.071]	3.001 (0.074) [0.071]	3.000 (0.075) [0.071]
	$\sigma^2 = 1$	0.931 (0.070) [0.067]	0.932 (0.162) [0.147]	0.930 (0.116) [0.106]
800	$\beta_1 = 1$	1.000 (0.049) [0.047]	0.999 (0.048) [0.047]	1.001 (0.050) [0.047]
	$\beta_2 = 2$	1.999 (0.048) [0.047]	1.999 (0.046) [0.047]	2.003 (0.048) [0.047]
	$\beta_3 = 3$	2.999 (0.052) [0.050]	3.000 (0.050) [0.050]	3.000 (0.051) [0.050]
	$\sigma^2 = 1$	0.965 (0.051) [0.049]	0.967 (0.118) [0.113]	0.966 (0.083) [0.082]
1200	$\beta_1 = 1$	1.000 (0.041) [0.039]	1.002 (0.040) [0.039]	0.999 (0.040) [0.039]
	$\beta_2 = 2$	2.000 (0.040) [0.039]	2.002 (0.039) [0.039]	2.000 (0.040) [0.039]
	$\beta_3 = 3$	2.999 (0.040) [0.040]	3.001 (0.040) [0.040]	3.000 (0.041) [0.040]
	$\sigma^2 = 1$	0.978 (0.041) [0.041]	0.975 (0.096) [0.094]	0.977 (0.068) [0.068]

Notes: Number of cluster and time segments are set to be fixed. `Queen Contiguity` for W_{l1} and W_{e1} and `Rook Contiguity` for W_{l2} and W_{e2}. Replication = 2000.

Table 2.2.A: Empirical Mean(sd)[$\hat{\text{se}}$] of QMLE and AQSE: STAR$(2,2)$ Model, $C = T = n/10$, Homoskedasticity

n	$\delta_0 = (\lambda_0, \rho_0)'$	Normal	Normal Mixture	Chi-Squared
		Spatiotemporal Effects: QMLE		
400	$\lambda_1 = 0.5$	0.499 (0.017) [0.015]	0.499 (0.017) [0.015]	0.498 (0.017) [0.015]
	$\lambda_2 = 0.3$	0.299 (0.016) [0.013]	0.299 (0.015) [0.013]	0.300 (0.016) [0.013]
	$\rho_1 = 0.2$	0.204 (0.104) [0.070]	0.200 (0.106) [0.071]	0.200 (0.109) [0.070]
	$\rho_2 = 0.4$	0.467 (0.073) [0.052]	0.464 (0.073) [0.052]	0.468 (0.073) [0.052]
800	$\lambda_1 = 0.5$	0.499 (0.014) [0.012]	0.499 (0.014) [0.012]	0.499 (0.014) [0.012]
	$\lambda_2 = 0.3$	0.299 (0.012) [0.010]	0.299 (0.012) [0.010]	0.299 (0.012) [0.010]
	$\rho_1 = 0.2$	0.222 (0.070) [0.050]	0.223 (0.070) [0.050]	0.220 (0.071) [0.050]
	$\rho_2 = 0.4$	0.473 (0.052) [0.037]	0.475 (0.050) [0.037]	0.474 (0.050) [0.037]
1200	$\lambda_1 = 0.5$	0.499 (0.011) [0.010]	0.499 (0.011) [0.010]	0.499 (0.011) [0.010]
	$\lambda_2 = 0.3$	0.300 (0.011) [0.009]	0.299 (0.011) [0.009]	0.299 (0.011) [0.009]
	$\rho_1 = 0.2$	0.222 (0.059) [0.041]	0.219 (0.059) [0.041]	0.225 (0.057) [0.041]
	$\rho_2 = 0.4$	0.475 (0.041) [0.030]	0.478 (0.040) [0.030]	0.475 (0.040) [0.030]
		Spatiotemporal Effects: AQSE		
400	$\lambda_1 = 0.5$	0.500 (0.017) [0.017]	0.500 (0.017) [0.016]	0.498 (0.017) [0.017]
	$\lambda_2 = 0.3$	0.299 (0.016) [0.016]	0.300 (0.014) [0.015]	0.300 (0.016) [0.016]
	$\rho_1 = 0.2$	0.194 (0.091) [0.095]	0.191 (0.095) [0.095]	0.192 (0.096) [0.095]
	$\rho_2 = 0.4$	0.394 (0.066) [0.070]	0.388 (0.075) [0.072]	0.392 (0.071) [0.070]
800	$\lambda_1 = 0.5$	0.499 (0.014) [0.014]	0.499 (0.014) [0.014]	0.499 (0.015) [0.014]
	$\lambda_2 = 0.3$	0.300 (0.012) [0.012]	0.300 (0.011) [0.011]	0.299 (0.012) [0.012]
	$\rho_1 = 0.2$	0.197 (0.062) [0.066]	0.194 (0.063) [0.066]	0.197 (0.064) [0.066]
	$\rho_2 = 0.4$	0.395 (0.048) [0.049]	0.396 (0.053) [0.049]	0.396 (0.048) [0.049]
1200	$\lambda_1 = 0.5$	0.499 (0.011) [0.011]	0.499 (0.011) [0.011]	0.500 (0.012) [0.011]
	$\lambda_2 = 0.3$	0.300 (0.011) [0.011]	0.300 (0.009) [0.009]	0.299 (0.011) [0.011]
	$\rho_1 = 0.2$	0.197 (0.053) [0.053]	0.196 (0.054) [0.053]	0.197 (0.053) [0.053]
	$\rho_2 = 0.4$	0.397 (0.037) [0.040]	0.397 (0.043) [0.040]	0.397 (0.040) [0.040]

Notes: Number of cluster and time segments increase. `Queen Contiguity` for W_{l1} and W_{e1} and `Rook Contiguity` for W_{l2} and W_{e2}. Replication = 2000.

Table 2.2.B: Empirical Mean(sd)[$\hat{\text{se}}$] of QMLE and AQSE: $\texttt{STAR}(2,2)$ Model, $C = T = n/10$, Homoskedasticity

n	$(\beta_0', \sigma_0^2)'$	Normal	Normal Mixture	Chi-Squared
		Covariate Coefficients and Error Variance: QMLE		
400	$\beta_1 = 1$	1.001 (0.082) [0.071]	1.001 (0.082) [0.071]	1.000 (0.081) [0.070]
	$\beta_2 = 2$	1.999 (0.080) [0.067]	2.001 (0.078) [0.067]	2.001 (0.077) [0.067]
	$\beta_3 = 3$	2.999 (0.081) [0.071]	2.996 (0.082) [0.071]	2.999 (0.081) [0.071]
	$\sigma^2 = 1$	0.762 (0.065) [0.055]	0.763 (0.134) [0.106]	0.760 (0.099) [0.079]
800	$\beta_1 = 1$	1.000 (0.057) [0.049]	0.998 (0.056) [0.049]	1.001 (0.056) [0.049]
	$\beta_2 = 2$	2.000 (0.055) [0.048]	2.001 (0.054) [0.047]	2.002 (0.053) [0.047]
	$\beta_3 = 3$	3.001 (0.059) [0.051]	3.002 (0.058) [0.051]	3.001 (0.057) [0.051]
	$\sigma^2 = 1$	0.768 (0.047) [0.040]	0.770 (0.096) [0.077]	0.767 (0.070) [0.058]
1200	$\beta_1 = 1$	1.000 (0.046) [0.040]	0.997 (0.045) [0.040]	1.001 (0.045) [0.040]
	$\beta_2 = 2$	2.001 (0.045) [0.039]	1.997 (0.044) [0.039]	2.000 (0.044) [0.039]
	$\beta_3 = 3$	2.999 (0.047) [0.041]	2.999 (0.046) [0.040]	2.999 (0.046) [0.040]
	$\sigma^2 = 1$	0.773 (0.038) [0.033]	0.770 (0.079) [0.064]	0.773 (0.060) [0.048]
		Covariate Coefficients and Error Variance: AQSE		
400	$\beta_1 = 1$	1.000 (0.082) [0.081]	1.000 (0.083) [0.081]	1.000 (0.080) [0.081]
	$\beta_2 = 2$	1.998 (0.079) [0.077]	2.000 (0.079) [0.077]	1.999 (0.078) [0.077]
	$\beta_3 = 3$	2.998 (0.081) [0.082]	3.000 (0.083) [0.081]	2.999 (0.082) [0.082]
	$\sigma^2 = 1$	0.975 (0.074) [0.081]	0.978 (0.161) [0.165]	0.975 (0.113) [0.121]
800	$\beta_1 = 1$	0.999 (0.057) [0.056]	0.998 (0.056) [0.056]	1.001 (0.056) [0.057]
	$\beta_2 = 2$	2.000 (0.055) [0.054]	1.998 (0.055) [0.056]	2.002 (0.055) [0.054]
	$\beta_3 = 3$	3.001 (0.059) [0.058]	2.999 (0.057) [0.057]	3.004 (0.058) [0.058]
	$\sigma^2 = 1$	0.988 (0.053) [0.058]	0.991 (0.114) [0.121]	0.989 (0.081) [0.089]
1200	$\beta_1 = 1$	1.000 (0.047) [0.045]	0.999 (0.046) [0.045]	1.000 (0.046) [0.045]
	$\beta_2 = 2$	2.000 (0.045) [0.045]	2.001 (0.047) [0.047]	1.999 (0.045) [0.045]
	$\beta_3 = 3$	2.999 (0.047) [0.046]	3.002 (0.047) [0.047]	2.999 (0.047) [0.046]
	$\sigma^2 = 1$	0.994 (0.043) [0.048]	0.990 (0.093) [0.099]	0.993 (0.067) [0.073]

Notes: Number of cluster and time segments increase. `Queen Contiguity` for W_{l1} and W_{e1} and `Rook Contiguity` for W_{l2} and W_{e2}. Replication = 2000.

Table 2.3.A: Empirical Mean(sd)[$\hat{se}$] of AQSE and RAQSE: $\texttt{STAR}(2,2)$ Model, $C = T = n/10$, Heteroskedasticity

n	$\delta_0 = (\lambda_0, \rho_0)'$	Normal	Normal Mixture	Chi-Squared
		Spatiotemporal Effects: AQSE		
400	$\lambda_1 = 0.5$	0.500 (0.013) [0.013]	0.500 (0.014) [0.012]	0.500 (0.014) [0.013]
	$\lambda_2 = 0.3$	0.301 (0.013) [0.012]	0.301 (0.013) [0.012]	0.300 (0.013) [0.012]
	$\rho_1 = 0.2$	0.182 (0.107) [0.102]	0.180 (0.098) [0.101]	0.182 (0.098) [0.101]
	$\rho_2 = 0.4$	0.386 (0.083) [0.086]	0.380 (0.088) [0.085]	0.382 (0.083) [0.084]
800	$\lambda_1 = 0.5$	0.501 (0.009) [0.011]	0.500 (0.010) [0.015]	0.500 (0.009) [0.011]
	$\lambda_2 = 0.3$	0.302 (0.010) [0.010]	0.301 (0.010) [0.013]	0.301 (0.010) [0.012]
	$\rho_1 = 0.2$	0.199 (0.087) [0.134]	0.194 (0.117) [1.172]	0.199 (0.072) [0.526]
	$\rho_2 = 0.4$	0.390 (0.079) [0.217]	0.400 (0.107) [0.462]	0.391 (0.071) [0.354]
1200	$\lambda_1 = 0.5$	0.500 (0.008) [0.016]	0.501 (0.009) [0.010]	0.500 (0.008) [0.011]
	$\lambda_2 = 0.3$	0.301 (0.008) [0.013]	0.301 (0.008) [0.010]	0.301 (0.008) [0.011]
	$\rho_1 = 0.2$	0.217 (0.126) [0.969]	0.203 (0.094) [0.505]	0.210 (0.121) [0.651]
	$\rho_2 = 0.4$	0.404 (0.061) [1.301]	0.397 (0.061) [0.681]	0.401 (0.057) [1.527]
		Spatiotemporal Effects: RAQSE		
400	$\lambda_1 = 0.5$	0.496 (0.037) [0.036]	0.500 (0.014) [0.013]	0.497 (0.036) [0.035]
	$\lambda_2 = 0.3$	0.300 (0.031) [0.030]	0.299 (0.013) [0.013]	0.300 (0.029) [0.030]
	$\rho_1 = 0.2$	0.196 (0.095) [0.097]	0.193 (0.098) [0.098]	0.195 (0.096) [0.095]
	$\rho_2 = 0.4$	0.393 (0.079) [0.080]	0.397 (0.080) [0.076]	0.395 (0.079) [0.079]
800	$\lambda_1 = 0.5$	0.499 (0.027) [0.027]	0.500 (0.009) [0.009]	0.498 (0.028) [0.027]
	$\lambda_2 = 0.3$	0.300 (0.023) [0.023]	0.300 (0.010) [0.010]	0.300 (0.024) [0.023]
	$\rho_1 = 0.2$	0.198 (0.067) [0.068]	0.196 (0.067) [0.066]	0.199 (0.067) [0.067]
	$\rho_2 = 0.4$	0.395 (0.057) [0.057]	0.397 (0.060) [0.056]	0.397 (0.057) [0.056]
1200	$\lambda_1 = 0.5$	0.499 (0.021) [0.021]	0.500 (0.009) [0.009]	0.499 (0.021) [0.021]
	$\lambda_2 = 0.3$	0.300 (0.018) [0.018]	0.300 (0.008) [0.008]	0.300 (0.018) [0.018]
	$\rho_1 = 0.2$	0.200 (0.055) [0.055]	0.201 (0.054) [0.054]	0.199 (0.053) [0.054]
	$\rho_2 = 0.4$	0.399 (0.046) [0.046]	0.399 (0.047) [0.046]	0.400 (0.045) [0.045]

Notes: Number of cluster and time segments increase. `Group Interaction` for W_{l1} and W_{e1} and `Rook Contiguity` for W_{l2} and W_{e2}. Replication = 2000.

Table 2.3.B: Empirical Mean(sd)[$\hat{\text{se}}$] of AQSE and RAQSE: $\texttt{STAR}(2,2)$ Model, $C = T = n/10$, Heteroskedasticity

n	$\delta_0 = (\lambda_0, \rho_0)'$	Normal	Normal Mixture	Chi-Squared
		Covariate Coefficients and Error Variance: AQSE		
400	$\beta_1 = 1$	1.001 (0.080) [0.078]	0.999 (0.078) [0.077]	0.999 (0.079) [0.078]
	$\beta_2 = 2$	1.998 (0.088) [0.087]	1.992 (0.091) [0.086]	2.001 (0.092) [0.086]
	$\beta_3 = 3$	3.002 (0.084) [0.079]	2.998 (0.085) [0.078]	3.000 (0.084) [0.079]
	$\sigma^2 = 1$	0.979 (0.105) [0.121]	0.972 (0.236) [0.227]	0.973 (0.169) [0.173]
800	$\beta_1 = 1$	0.997 (0.056) [0.064]	0.996 (0.059) [0.107]	1.000 (0.055) [0.068]
	$\beta_2 = 2$	2.000 (0.054) [0.063]	2.001 (0.057) [0.103]	1.999 (0.055) [0.062]
	$\beta_3 = 3$	2.999 (0.054) [0.058]	2.998 (0.055) [0.099]	2.999 (0.052) [0.065]
	$\sigma^2 = 1$	0.988 (0.075) [0.124]	0.987 (0.163) [0.254]	0.983 (0.115) [0.137]
1200	$\beta_1 = 1$	1.000 (0.045) [0.092]	1.000 (0.047) [0.066]	0.999 (0.044) [0.069]
	$\beta_2 = 2$	1.998 (0.044) [0.066]	1.999 (0.049) [0.064]	2.000 (0.045) [0.108]
	$\beta_3 = 3$	3.002 (0.045) [0.084]	2.999 (0.045) [0.063]	3.002 (0.046) [0.088]
	$\sigma^2 = 1$	0.994 (0.060) [0.250]	0.990 (0.136) [0.256]	0.992 (0.099) [0.418]
		Covariate Coefficients: RAQSE		
400	$\beta_1 = 1$	0.994 (0.299) [0.294]	1.001 (0.078) [0.077]	0.998 (0.282) [0.291]
	$\beta_2 = 2$	1.996 (0.325) [0.323]	2.001 (0.087) [0.086]	2.003 (0.325) [0.321]
	$\beta_3 = 3$	3.010 (0.290) [0.291]	3.002 (0.085) [0.083]	2.994 (0.288) [0.290]
800	$\beta_1 = 1$	0.993 (0.225) [0.224]	0.998 (0.057) [0.055]	1.006 (0.213) [0.210]
	$\beta_2 = 2$	2.001 (0.216) [0.217]	2.000 (0.054) [0.054]	1.992 (0.217) [0.215]
	$\beta_3 = 3$	3.004 (0.215) [0.213]	3.000 (0.055) [0.053]	3.000 (0.214) [0.214]
1200	$\beta_1 = 1$	0.998 (0.170) [0.168]	1.000 (0.043) [0.043]	0.999 (0.169) [0.174]
	$\beta_2 = 2$	1.996 (0.166) [0.172]	2.000 (0.044) [0.044]	1.997 (0.175) [0.173]
	$\beta_3 = 3$	3.002 (0.180) [0.179]	3.000 (0.047) [0.046]	2.998 (0.177) [0.177]

Notes: Number of cluster and time segments increase. `Group Interaction` for W_{l1} and W_{e1} and `Rook Contiguity` for W_{l2} and W_{e2}. Replication = 2000.

Table 2.4: Variables and Summary Statistics

(1) Variable	(2) Mean	(3) SD	(4) Min	(5) Max
Total price (10,000 RMB)	296.77	173.19	10.00	4370.00
Living area (square meters)	84.72	35.17	10.00	460.24
Age	15.51	8.66	-1	65
Number of bedrooms	2.04	0.75	0	8
Number of living rooms	1.19	0.52	0	5
Number of kitchens	1.00	0.08	0	3
Number of bathrooms	1.19	0.42	0	7
$\mathbb{1}$(Ground Floor)	8.1%	0.27	0	1
$\mathbb{1}$(Low Floor)	19.8%	0.40	0	1
$\mathbb{1}$(Middle Floor)	38.1%	0.49	0	1
$\mathbb{1}$(High Floor)	21.8%	0.41	0	1
$\mathbb{1}$(Top Floor)	11.8%	0.32	0	1
$\mathbb{1}$(Tower-type)	23.6%	0.42	0	1
$\mathbb{1}$(Slab-type)	57.5%	0.49	0	1
$\mathbb{1}$(Tower-slab Combined)	18.6%	0.39	0	1
$\mathbb{1}$(Rough House)	2.6%	0.16	0	1
$\mathbb{1}$(Simple Decoration)	33.3%	0.47	0	1
$\mathbb{1}$(Fine Decoration)	51.5%	0.50	0	1
$\mathbb{1}$(Other)	12.6%	0.33	0	1
$\mathbb{1}$(Brick-concrete Structure)	1.6%	0.12	0	1
$\mathbb{1}$(Steel-concrete Structure)	57.1%	0.49	0	1
$\mathbb{1}$(Mixed Structure)	41.2%	0.49	0	1
$\mathbb{1}$(Elevator)	57.0%	0.50	0	1

Notes: Tower-type and slab-type are two major tall-building types for apartments in China. For tower-type buildings, there are usually more than four or five housing units surrounding the elevators placed in the center of a building, thus units are usually not north-south transparent. Slab-type buildings are in shape of rectangles and usually north-south transparent. Typically, apartments in slab-type buildings have higher prices.

Table 2.5: Estimation Results of the $\texttt{STAR}(1,0)$, $\texttt{STAR}(2,0)$ and $\texttt{STAR}(2,2)$ Models

Model	Sample	λ_{cmty} QMLE (1)	λ_{cmty} AQSE (2)	λ_{cmty} RAQSE (3)	λ_{nbr} QMLE (4)	λ_{nbr} AQSE (5)	λ_{nbr} RAQSE (6)
$\texttt{STAR}(1,0)$	Mar-Apr	0.016140	0.030085	0.013593	0.255369	0.254184	0.424507
		(0.001185)	(0.001713)	(0.002267)	(0.005048)	(0.007884)	(0.018394)
	May-Jun	0.021735	0.029397	0.012139	0.299928	0.277599	0.480418
		(0.001218)	(0.001785)	(0.002314)	(0.005178)	(0.010033)	(0.016619)
	Jul-Aug	0.014790	0.024073	0.007465	0.305316	0.279621	0.504543
		(0.001267)	(0.001815)	(0.002376)	(0.005514)	(0.010244)	(0.017119)
	Sep-Oct	0.016604	0.021193	0.007946	0.302217	0.287274	0.503541
		(0.001231)	(0.001876)	(0.002400)	(0.005760)	(0.011207)	(0.015403)
	Nov-Dec	0.027112	0.041902	0.018454	0.297106	0.294556	0.498481
		(0.001244)	(0.001758)	(0.002550)	(0.004474)	(0.008558)	(0.012669)
$\texttt{STAR}(2,0)$	Mar-Apr	0.011237	0.022579	0.008611	0.251363	0.245867	0.41715
		(0.001103)	(0.001297)	(0.001540)	(0.005034)	(0.008149)	(0.018107)
	May-Jun	0.015589	0.025389	0.010673	0.293912	0.268285	0.471669
		(0.001113)	(0.001357)	(0.001485)	(0.005145)	(0.010379)	(0.016391)
	Jul-Aug	0.010400	0.019965	0.006877	0.302333	0.273531	0.499347
		(0.001157)	(0.001339)	(0.001538)	(0.005500)	(0.010504)	(0.01701)
	Sep-Oct	0.011224	0.017827	0.006454	0.297600	0.280702	0.497848
		(0.001125)	(0.001344)	(0.001489)	(0.005753)	(0.011490)	(0.015227)
	Nov-Dec	0.019270	0.034716	0.014624	0.290305	0.280418	0.48675
		(0.001131)	(0.001372)	(0.001660)	(0.004442)	(0.009001)	(0.012521)
$\texttt{STAR}(2,2)$	Mar-Apr	-0.004653	0.017244	-0.002205	0.002437	0.124535	0.013425
		(0.001037)	(0.001610)	(0.000968)	(0.006205)	(0.020092)	(0.021115)
	May-Jun	0.001415	0.019328	0.000538	0.147218	0.138294	-0.02137
		(0.001063)	(0.002015)	(0.001079)	(0.006550)	(0.020068)	(0.03076)
	Jul-Aug	0.015776	0.013718	-0.000701	0.213803	0.136857	-0.03741
		(0.001113)	(0.001888)	(0.001368)	(0.007166)	(0.020553)	(0.025788)
	Sep-Oct	0.000994	0.013345	-0.000473	-0.016390	0.149845	-0.05325
		(0.001193)	(0.001707)	(0.001207)	(0.007913)	(0.022015)	(0.034732)
	Nov-Dec	0.000325	0.027002	0.002953	0.100034	0.136576	-0.01241
		(0.000996)	(0.002327)	(0.001685)	(0.005914)	(0.017337)	(0.031196)

Notes: This table reports the coefficients of the spatiotemporal interaction effects estimated using $\texttt{STAR}(1,0)$, $\texttt{STAR}(2,0)$ and $\texttt{STAR}(2,2)$ Models. Standard errors are reported in parentheses.

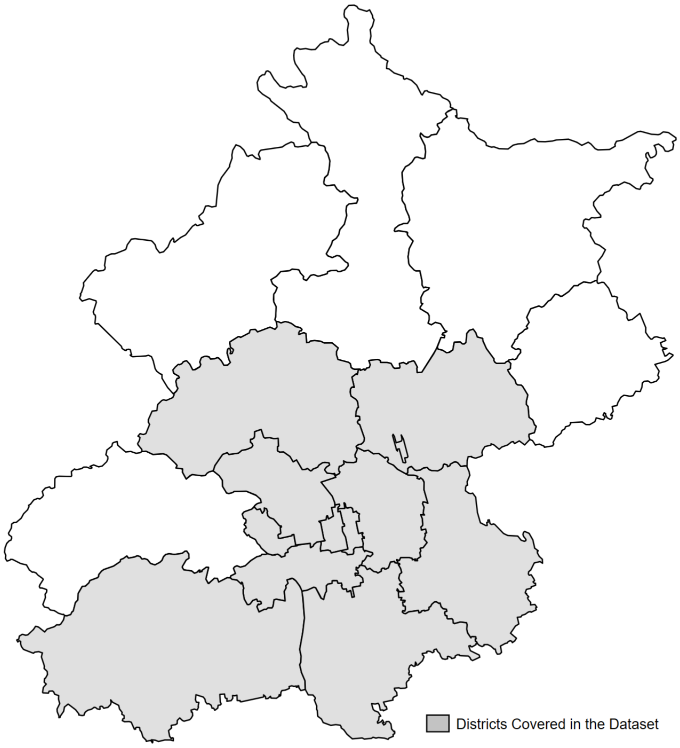


Figure 2.1: Map of Municipal Districts of Beijing

Notes: The shadow areas are the municipal districts covered in the dataset. These districts are more urbanized than the peripheral districts in Beijing.

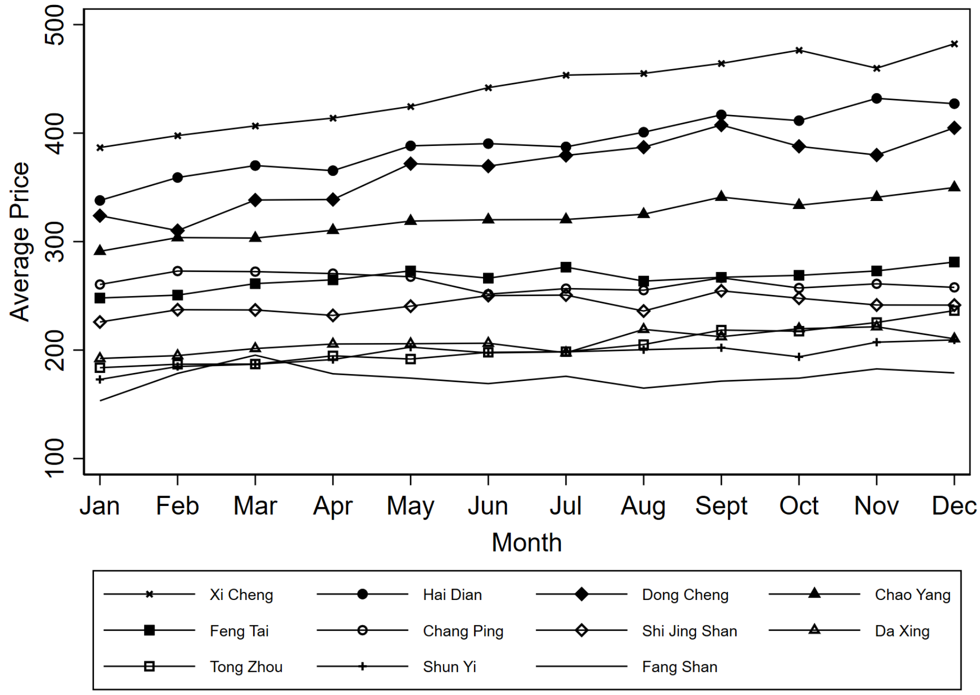


Figure 2.2: Average Price by Month for Different Districts

Notes: This figure illustrates the trends of the average housing transaction prices for different districts in Beijing. The prices are counted in 10,000 RMB.

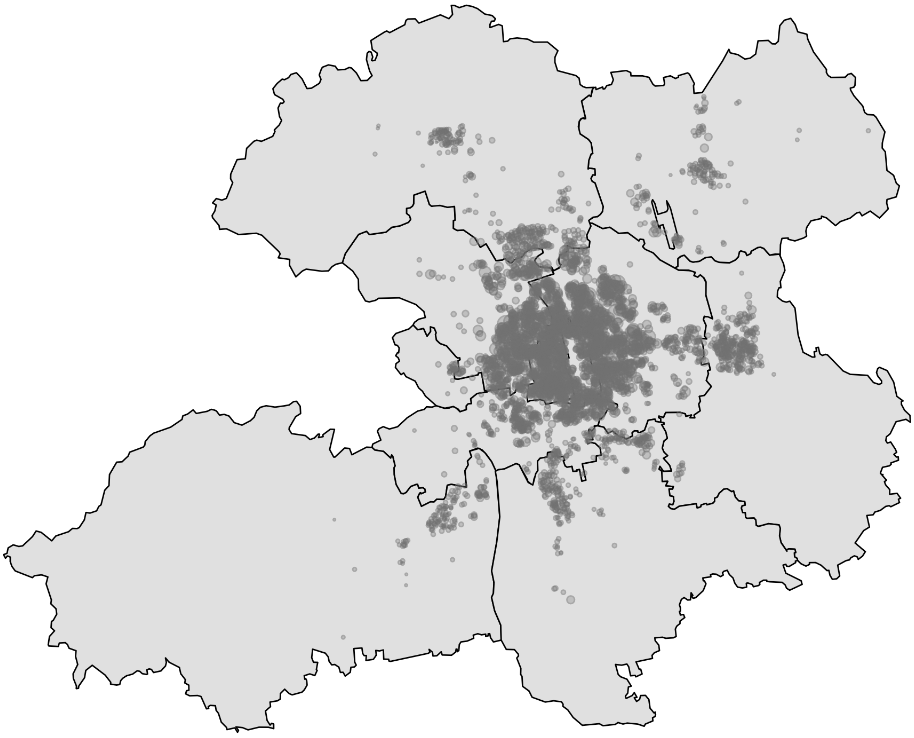


Figure 2.3: Community Level Average Price

Notes: Each bubble in the figure is a community in Beijing. Larger bubbles represents higher transaction prices.

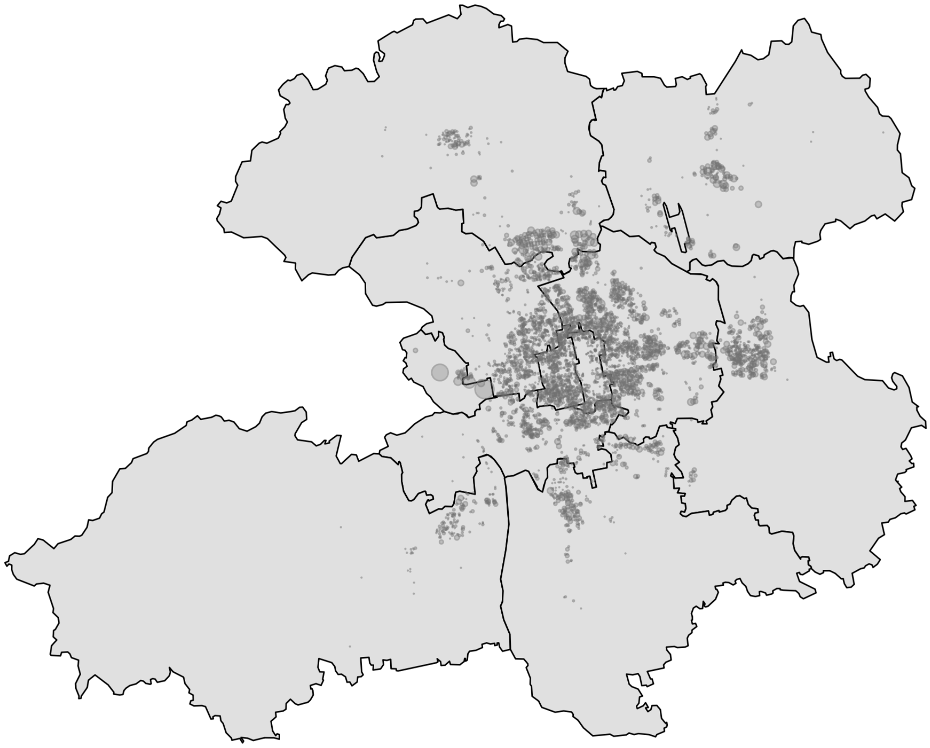


Figure 2.4: Locations of Housing Transactions by Communities

Notes: Each bubble is a community in Beijing. Larger bubbles represents more housing transactions in a community. We define the $W_{d,cmty}$ in Section 2.6 based on transactions in the same communities.

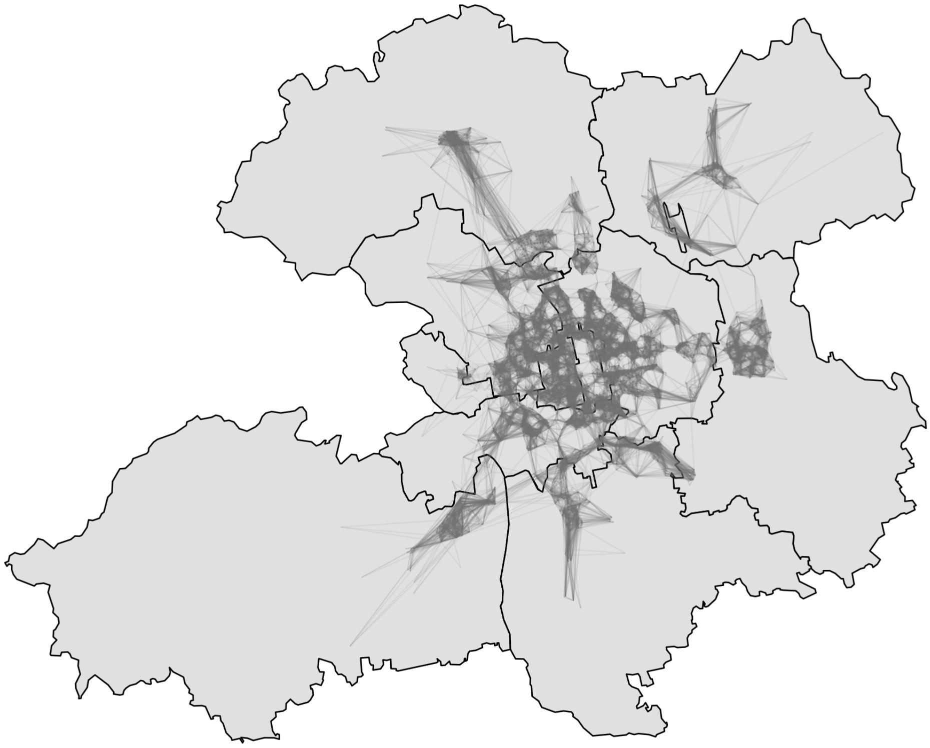


Figure 2.5: 5 Nearest Neighbours outside of Communities

Notes: This figure shows the connectivity of a housing transaction to its 5 nearest neighbours (5NN) one month before or after the transaction takes place.

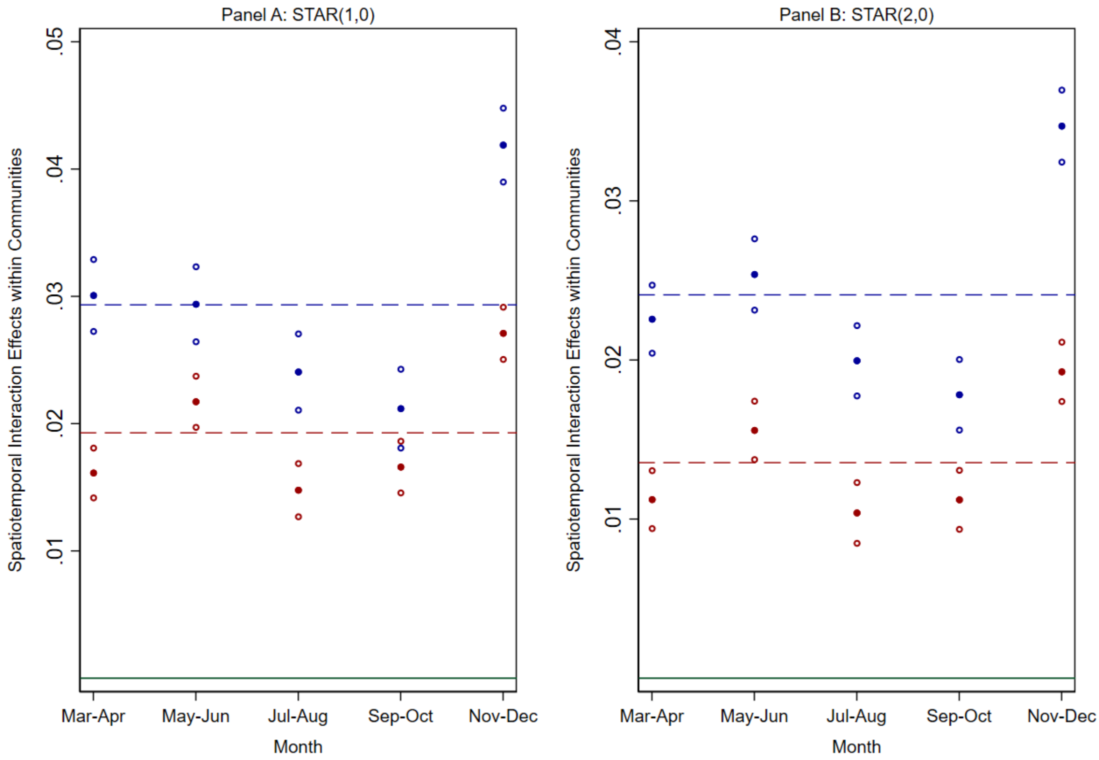


Figure 2.6.A: Spatiotemporal Interaction Effects within Communities: QMLE and AQSE

Notes: The solid points are estimates of the spatiotemporal interaction effects within communities. The hollow points represent the 90% confidence intervals. The QML estimates (no community fixed effects) are in red, while the AQS estimates (allow for community fixed effects) are in blue. The dash lines are the average of the estimates.

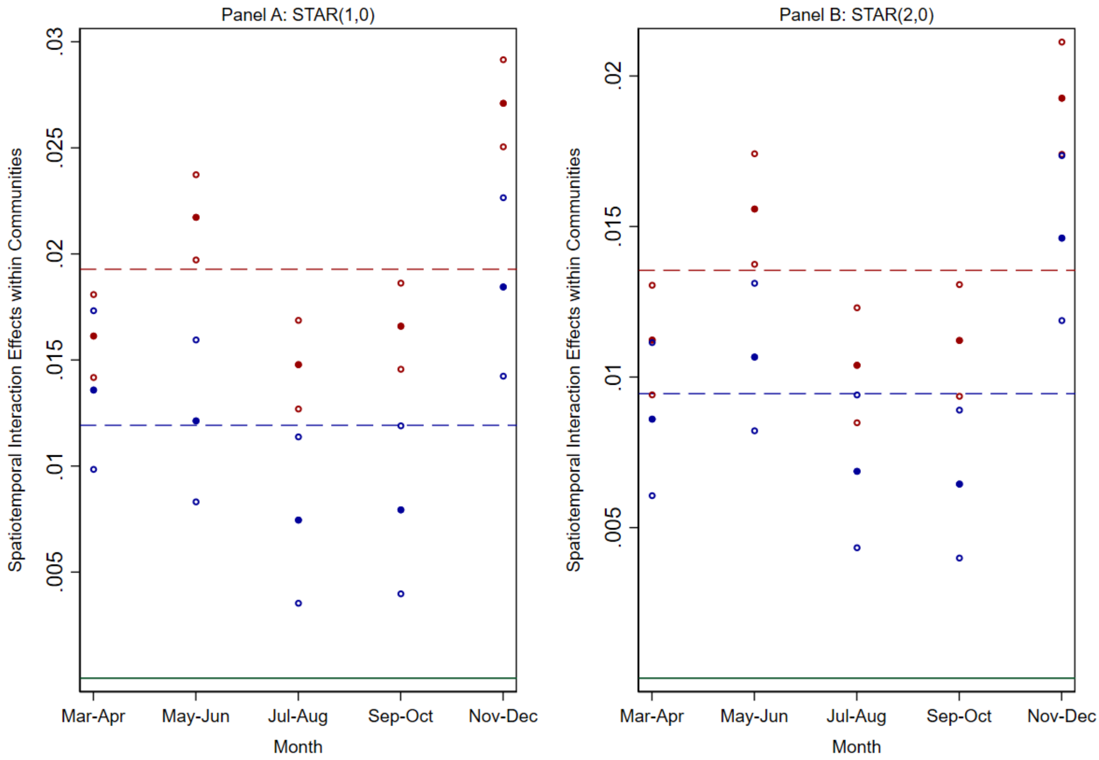


Figure 2.6.B: Spatiotemporal Interaction Effects within Communities: QMLE and RAQSE

Notes: The solid points are estimates of the spatiotemporal interaction effects within communities. The hollow points represent the 90% confidence intervals. The QML estimates (no community fixed effects) are in red, while the RAQS estimates (allow for community fixed effects and heteroskedasticity) are in blue. The dash lines are the average of the estimates.

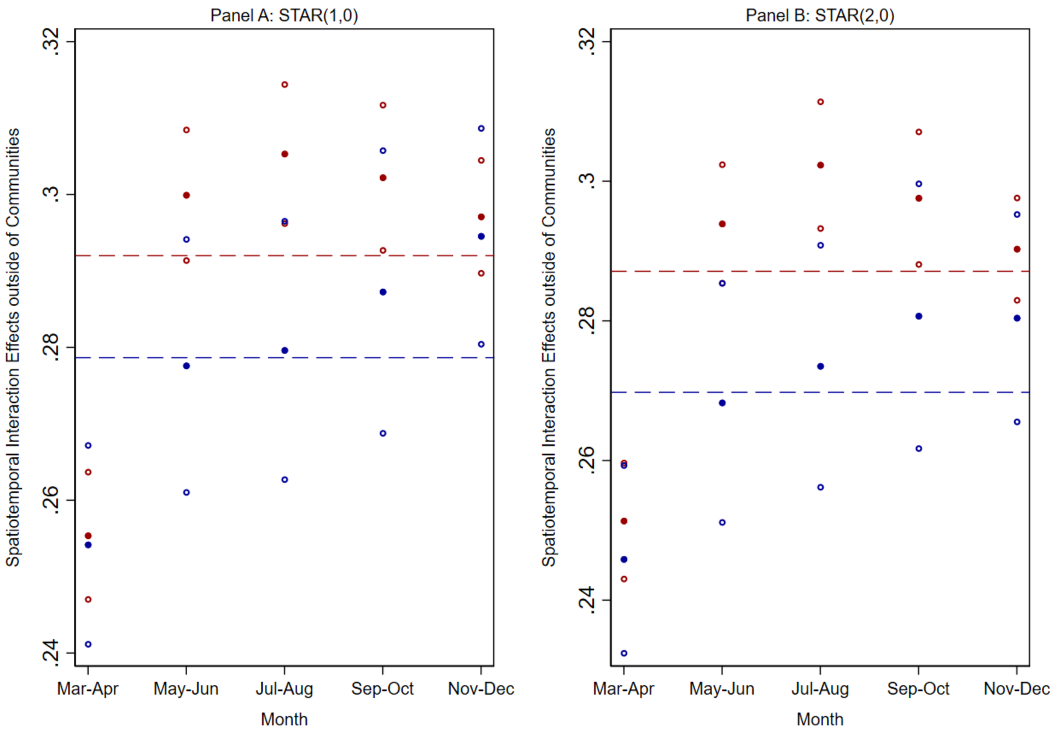


Figure 2.7.A: Effects of 5 Nearest Neighbours outside of Communities: QMLE and AQSE

Notes: The solid points are estimates of the spatiotemporal interaction effects of 5 nearest neighbours outside of communities. The hollow points represent the 90% confidence intervals. The QML estimates (no community fixed effects) are in red, while the RAQS estimates (allow for community fixed effects) are in blue. The dash lines are the average of the estimates.

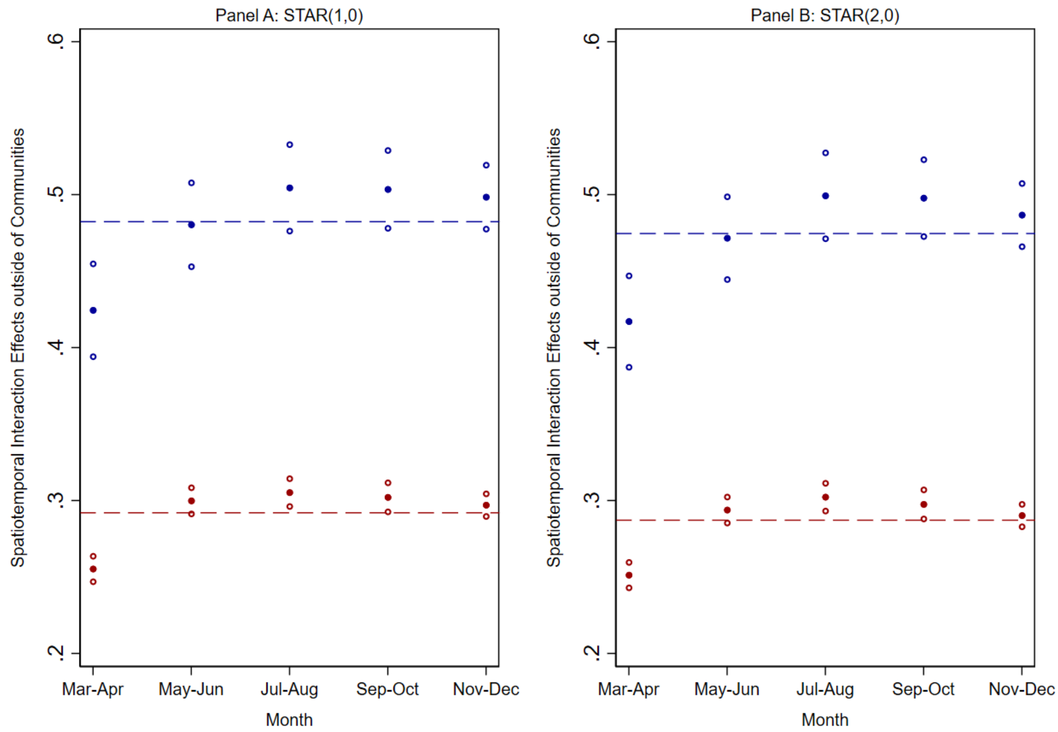


Figure 2.7.B: Effects of 5 Nearest Neighbours outside of Communities: QMLE and RAQSE

Notes: The solid points are estimates of the spatiotemporal interaction effects of 5 nearest neighbours outside of communities. The hollow points represent the 90% confidence intervals. The QML estimates (no community fixed effects) are in red, while the RAQS estimates (allow for community fixed effects and heteroskedasticity) are in blue. The dash lines are the average of the estimates.

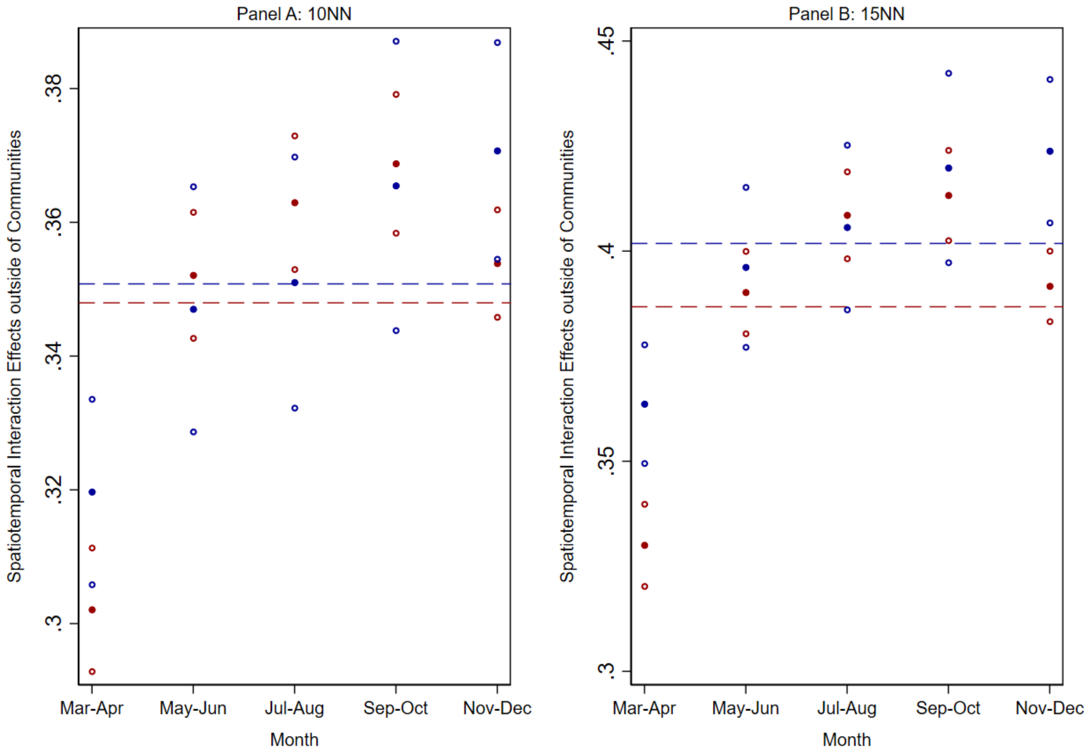


Figure 2.8: Effects of 10 and 15 Nearest Neighbours outside of Communities: QMLE and AQSE

Notes: The solid points are estimates of the spatiotemporal interaction effects of 10 and 15 nearest neighbours outside of communities. The hollow points represent the 90% confidence intervals. The QML estimates (no community fixed effects) are in red, while the AQS estimates (allow for community fixed effects) are in blue. The dash lines are the average of the estimates.

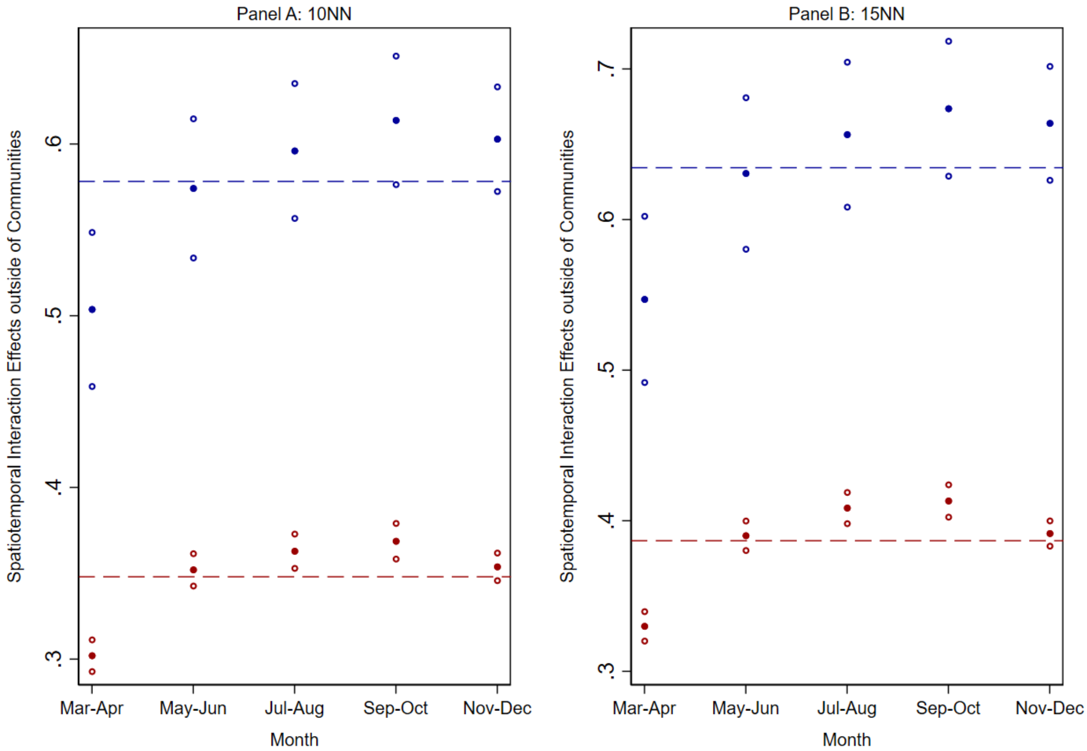


Figure 2.9: Effects of 10 and 15 Nearest Neighbours outside of Communities: QMLE and RAQSE

Notes: The solid points are estimates of the spatiotemporal interaction effects of 10 nearest neighbours outside of communities. The hollow points represent the 90% confidence intervals. The QML estimates (no community fixed effects) are in red, while the RAQS estimates (allow for community fixed effects and heteroskedasticity) are in blue. The dash lines are the average of the estimates.

Table 2.6: Coefficients of Housing Attributes Obtained from STAR$(1,0)$ Model

Sample	Variables	QMLE Coefficients (1)	QMLE SE (2)	AQSE Coefficients (3)	AQSE SE (4)	RAQSE Coefficients (5)	RAQSE SE (6)
Mar-April	λ_{cmty}	0.016140	(0.00119)	0.030085	(0.00171)	0.013593	(0.00227)
	Living area (square meters)	0.013556	(0.00024)	0.013606	(0.00030)	0.013640	(0.00054)
	Living area^2	-2.5E-05	(8.3E-07)	-2.5E-05	(1.0E-06)	-2.5E-05	(2.3E-06)
	Age	0.002745	(0.00030)	0.010394	(0.00038)	0.010078	(0.00049)
	Number of bedrooms	0.042552	(0.00397)	0.050244	(0.00486)	0.049366	(0.00560)
	Number of living rooms	0.060587	(0.00462)	0.031546	(0.00559)	0.033277	(0.00561)
	Number of kitchens	0.155386	(0.02503)	0.185019	(0.02908)	0.184932	(0.06025)
	Number of bathrooms	-0.00775	(0.00637)	-0.00267	(0.00771)	-0.00142	(0.00936)
May-Jun	λ_{cmty}	0.021735	(0.00122)	0.029397	(0.00179)	0.012139	(0.00231)
	Living area (square meters)	0.013926	(0.00022)	0.013924	(0.00029)	0.013862	(0.00055)
	Living area^2	-2.5E-05	(7.6E-07)	-2.3E-05	(9.4E-07)	-2.3E-05	(2.2E-06)
	Age	0.003237	(0.00030)	0.012872	(0.00039)	0.012379	(0.00049)
	Number of bedrooms	0.037893	(0.00399)	0.024583	(0.00495)	0.024422	(0.00580)
	Number of living rooms	0.057006	(0.00444)	0.034071	(0.00544)	0.035439	(0.00559)
	Number of kitchens	0.233417	(0.02239)	0.256634	(0.02642)	0.259964	(0.07168)
	Number of bathrooms	-0.00223	(0.00605)	-0.00251	(0.00741)	-0.00103	(0.00964)
Jul-Aug	λ_{cmty}	0.014789	(0.00127)	0.024073	(0.00182)	0.007465	(0.00238)
	Living area (square meters)	0.012624	(0.00023)	0.012569	(0.00029)	0.012557	(0.00049)
	Living area^2	-2.1E-05	(7.5E-07)	-2.0E-05	(9.0E-07)	-2.0E-05	(1.9E-06)
	Age	0.002799	(0.00033)	0.013577	(0.00042)	0.013108	(0.00051)
	Number of bedrooms	0.049736	(0.00421)	0.040871	(0.00517)	0.040600	(0.00587)
	Number of living rooms	0.067692	(0.00473)	0.045072	(0.00570)	0.047075	(0.00592)
	Number of kitchens	0.229921	(0.02327)	0.279473	(0.02706)	0.283940	(0.05622)
	Number of bathrooms	-0.00667	(0.00641)	-0.00190	(0.00775)	-0.00074	(0.00933)
Sep-Oct	λ_{cmty}	0.016604	(0.00123)	0.021193	(0.00188)	0.007946	(0.00240)
	Living area (square meters)	0.012574	(0.00025)	0.01193	(0.00032)	0.011975	(0.00058)
	Living area^2	-2.1E-05	(8.1E-07)	-1.8E-05	(1.0E-06)	-1.8E-05	(2.3E-06)
	Age	0.001739	(0.00034)	0.012457	(0.00044)	0.012099	(0.00053)
	Number of bedrooms	0.060072	(0.00440)	0.059404	(0.00565)	0.058888	(0.00668)
	Number of living rooms	0.069108	(0.00501)	0.047100	(0.00627)	0.047890	(0.00666)
	Number of kitchens	0.115660	(0.02369)	0.088895	(0.02890)	0.090886	(0.05187)
	Number of bathrooms	-0.02247	(0.00695)	-0.00963	(0.00872)	-0.00885	(0.01007)
Nov-Dec	λ_{cmty}	0.027112	(0.00124)	0.041902	(0.00176)	0.018454	(0.00255)
	Living area (square meters)	0.012695	(0.00019)	0.012823	(0.00024)	0.012851	(0.00046)
	Living area^2	-2.2E-05	(6.4E-07)	-2.1E-05	(7.6E-07)	-2.1E-05	(1.8E-06)
	Age	0.001460	(0.00028)	0.012470	(0.00036)	0.011984	(0.00046)
	Number of bedrooms	0.057713	(0.00351)	0.046227	(0.00426)	0.045648	(0.00507)
	Number of living rooms	0.068831	(0.00392)	0.038626	(0.00467)	0.039813	(0.00489)
	Number of kitchens	0.143363	(0.01939)	0.138595	(0.02236)	0.138510	(0.03367)
	Number of bathrooms	-0.02011	(0.00551)	-0.01228	(0.00660)	-0.01059	(0.00765)
Other Control Variables		Yes		Yes		Yes	
District Fixed Effects		Yes		No		No	
Community Fixed Effects		No		Yes		Yes	
Week Fixed Effects		Yes		Yes		Yes	

Notes: This table reports the coefficients of housing attributes obtained from the STAR$(1,0)$ specification of (2.32) and (2.33).. Other control variables include indicator variables for floors, building types, decoration condition, building structures, and elevator. Standard errors are reported in parentheses.

Table 2.7: Coefficients of Housing Attributes Obtained from STAR$(1,0)$ Model

Sample	Variables	QMLE		AQSE		RAQSE	
		Coefficients	SE	Coefficients	SE	Coefficients	SE
		(1)	(2)	(3)	(4)	(5)	(6)
Mar-April	λ_{nbr}	0.255369	(0.00505)	0.254184	(0.00788)	0.424510	(0.01839)
	Living area (square meters)	0.013886	(0.00022)	0.013774	(0.00026)	0.013844	(0.00049)
	Living area^2	-2.6E-05	(7.7E-07)	-2.6E-05	(8.8E-07)	-2.7E-05	(2.2E-06)
	Age	0.002529	(0.00028)	0.007957	(0.00034)	0.006710	(0.00040)
	Number of bedrooms	0.042031	(0.00369)	0.050634	(0.00423)	0.051969	(0.00468)
	Number of living rooms	0.051684	(0.00430)	0.031200	(0.00487)	0.028853	(0.00494)
	Number of kitchens	0.166888	(0.02328)	0.179775	(0.02534)	0.176370	(0.05553)
	Number of bathrooms	-0.01544	(0.00592)	-0.00876	(0.00672)	-0.01436	(0.00827)
May-Jun	λ_{nbr}	0.299928	(0.00518)	0.277599	(0.01003)	0.480420	(0.01662)
	Living area (square meters)	0.013904	(0.00021)	0.013929	(0.00025)	0.014010	(0.00052)
	Living area^2	-2.5E-05	(7.0E-07)	-2.4E-05	(8.0E-07)	-2.5E-05	(2.2E-06)
	Age	0.003304	(0.00028)	0.009884	(0.00035)	0.008314	(0.00050)
	Number of bedrooms	0.044993	(0.00366)	0.032442	(0.00423)	0.038384	(0.00514)
	Number of living rooms	0.049153	(0.00406)	0.033177	(0.00464)	0.030821	(0.00453)
	Number of kitchens	0.257202	(0.02047)	0.262958	(0.02255)	0.263430	(0.07452)
	Number of bathrooms	-0.01648	(0.00553)	-0.01027	(0.00633)	-0.01779	(0.00883)
Jul-Aug	λ_{nbr}	0.305316	(0.00551)	0.279621	(0.01024)	0.504540	(0.01712)
	Living area (square meters)	0.012542	(0.00021)	0.012505	(0.00025)	0.012467	(0.00046)
	Living area^2	-2.1E-05	(6.9E-07)	-2.1E-05	(7.7E-07)	-2.1E-05	(1.9E-06)
	Age	0.002116	(0.00030)	0.010094	(0.00038)	0.007840	(0.00042)
	Number of bedrooms	0.056805	(0.00385)	0.047979	(0.00444)	0.054014	(0.00496)
	Number of living rooms	0.062068	(0.00432)	0.045507	(0.00488)	0.043522	(0.00503)
	Number of kitchens	0.244876	(0.02124)	0.280156	(0.02317)	0.275500	(0.05243)
	Number of bathrooms	-0.01654	(0.00585)	-0.00640	(0.00664)	-0.01137	(0.00839)
Sep-Oct	λ_{nbr}	0.302217	(0.00576)	0.287274	(0.01121)	0.503540	(0.01540)
	Living area (square meters)	0.012473	(0.00023)	0.011991	(0.00027)	0.011983	(0.00060)
	Living area^2	-2.1E-05	(7.4E-07)	-1.9E-05	(8.6E-07)	-1.9E-05	(2.5E-06)
	Age	0.001820	(0.00031)	0.009236	(0.00039)	0.007242	(0.00048)
	Number of bedrooms	0.065956	(0.00402)	0.063545	(0.00477)	0.067284	(0.00568)
	Number of living rooms	0.057970	(0.00458)	0.044087	(0.00529)	0.040867	(0.00561)
	Number of kitchens	0.129663	(0.02163)	0.099862	(0.02441)	0.105720	(0.04844)
	Number of bathrooms	-0.03451	(0.00635)	-0.01756	(0.00738)	-0.02447	(0.00871)
Nov-Dec	λ_{nbr}	0.297106	(0.00447)	0.294556	(0.00856)	0.498480	(0.01267)
	Living area (square meters)	0.012866	(0.00018)	0.012921	(0.00021)	0.012953	(0.00044)
	Living area^2	-2.3E-05	(5.8E-07)	-2.2E-05	(6.5E-07)	-2.3E-05	(1.8E-06)
	Age	0.001432	(0.00025)	0.009254	(0.00032)	0.007629	(0.00037)
	Number of bedrooms	0.058071	(0.00319)	0.048328	(0.00362)	0.050498	(0.00428)
	Number of living rooms	0.059344	(0.00357)	0.037316	(0.00396)	0.034941	(0.00409)
	Number of kitchens	0.166940	(0.01765)	0.142462	(0.01897)	0.145250	(0.03333)
	Number of bathrooms	-0.02625	(0.00501)	-0.01380	(0.00560)	-0.01695	(0.00683)
Other Control Variables		Yes		Yes		Yes	
District Fixed Effects		Yes		No		No	
Community Fixed Effects		No		Yes		Yes	
Week Fixed Effects		Yes		Yes		Yes	

Notes: This table reports the coefficients of housing attributes of the STAR$(2,2)$ specification of (2.32) and (2.33). Other control variables include indicator variables for floors, building types, decoration condition, building structures, and elevator. Standard errors are reported in parentheses.

Table 2.8: Coefficients of Housing Attributes Obtained from STAR$(2,0)$ Model

Sample	Variables	QMLE		AQSE		RAQSE	
		Coefficients	SE	Coefficients	SE	Coefficients	SE
		(1)	(2)	(3)	(4)	(5)	(6)
Mar-April	λ_{cmty}	0.011237	(0.00110)	0.022578	(0.00130)	0.008611	(0.00154)
	λ_{nbr}	0.251363	(0.00503)	0.245867	(0.00815)	0.417150	(0.01811)
	Living area (square meters)	0.013826	(0.00022)	0.013671	(0.00026)	0.013803	(0.00050)
	Living area^2	-2.6E-05	(7.7E-07)	-2.6E-05	(8.8E-07)	-2.7E-05	(2.2E-06)
	Age	0.002691	(0.00028)	0.008358	(0.00035)	0.006893	(0.00041)
	Number of bedrooms	0.042481	(0.00368)	0.052080	(0.00424)	0.052488	(0.00469)
	Number of living rooms	0.050494	(0.00429)	0.028967	(0.00488)	0.028059	(0.00493)
	Number of kitchens	0.166349	(0.02322)	0.179381	(0.02540)	0.176300	(0.05571)
	Number of bathrooms	-0.01622	(0.00591)	-0.01046	(0.00674)	-0.01487	(0.00829)
May-Jun	λ_{cmty}	0.015589	(0.00111)	0.025389	(0.00136)	0.010673	(0.00148)
	λ_{nbr}	0.293912	(0.00515)	0.268285	(0.01038)	0.471670	(0.01639)
	Living area (square meters)	0.013891	(0.00021)	0.013975	(0.00025)	0.014028	(0.00052)
	Living area^2	-2.5E-05	(6.9E-07)	-2.4E-05	(8.0E-07)	-2.5E-05	(2.1E-06)
	Age	0.003611	(0.00028)	0.010583	(0.00036)	0.008646	(0.00050)
	Number of bedrooms	0.045398	(0.00364)	0.033208	(0.00424)	0.038565	(0.00514)
	Number of living rooms	0.048020	(0.00404)	0.031647	(0.00465)	0.030234	(0.00452)
	Number of kitchens	0.252282	(0.02037)	0.256731	(0.02259)	0.260810	(0.07452)
	Number of bathrooms	-0.01719	(0.00551)	-0.01203	(0.00634)	-0.01835	(0.00881)
Jul-Aug	λ_{cmty}	0.010400	(0.00116)	0.019965	(0.00134)	0.006877	(0.00154)
	λ_{nbr}	0.302333	(0.00550)	0.273531	(0.01050)	0.499350	(0.01701)
	Living area (square meters)	0.012510	(0.00021)	0.012481	(0.00025)	0.012459	(0.00046)
	Living area^2	-2.1E-05	(6.9E-07)	-2.0E-05	(7.7E-07)	-2.1E-05	(1.9E-06)
	Age	0.002333	(0.00030)	0.010658	(0.00039)	0.008065	(0.00043)
	Number of bedrooms	0.057020	(0.00385)	0.048531	(0.00445)	0.054120	(0.00496)
	Number of living rooms	0.060904	(0.00431)	0.043461	(0.00489)	0.042845	(0.00502)
	Number of kitchens	0.242895	(0.02119)	0.275676	(0.02324)	0.274020	(0.05289)
	Number of bathrooms	-0.01698	(0.00584)	-0.00767	(0.00666)	-0.01174	(0.00842)
Sep-Oct	λ_{cmty}	0.011224	(0.00112)	0.017827	(0.00134)	0.006454	(0.00149)
	λ_{nbr}	0.297600	(0.00575)	0.280702	(0.01149)	0.497850	(0.01523)
	Living area (square meters)	0.012389	(0.00023)	0.011869	(0.00027)	0.011938	(0.00061)
	Living area^2	-2.0E-05	(7.4E-07)	-1.8E-05	(8.6E-07)	-1.9E-05	(2.5E-06)
	Age	0.001982	(0.00031)	0.009635	(0.00040)	0.007417	(0.00048)
	Number of bedrooms	0.066683	(0.00401)	0.064868	(0.00478)	0.067705	(0.00569)
	Number of living rooms	0.057010	(0.00456)	0.042652	(0.00530)	0.040397	(0.00562)
	Number of kitchens	0.128645	(0.02156)	0.099201	(0.02446)	0.105390	(0.04859)
	Number of bathrooms	-0.03464	(0.00633)	-0.01830	(0.00739)	-0.02463	(0.00873)
Nov-Dec	λ_{cmty}	0.019270	(0.00113)	0.034716	(0.00137)	0.014624	(0.00166)
	λ_{nbr}	0.290305	(0.00444)	0.280418	(0.00900)	0.486750	(0.01252)
	Living area (square meters)	0.012764	(0.00018)	0.012817	(0.00021)	0.012909	(0.00045)
	Living area^2	-2.3E-05	(5.8E-07)	-2.1E-05	(6.5E-07)	-2.2E-05	(1.8E-06)
	Age	0.001705	(0.00025)	0.009891	(0.00032)	0.007944	(0.00037)
	Number of bedrooms	0.058756	(0.00317)	0.049642	(0.00363)	0.050991	(0.00426)
	Number of living rooms	0.058190	(0.00354)	0.036443	(0.00397)	0.034640	(0.00408)
	Number of kitchens	0.164822	(0.01754)	0.141038	(0.01902)	0.144570	(0.03346)
	Number of bathrooms	-0.02731	(0.00498)	-0.01653	(0.00561)	-0.01801	(0.00682)
Other Control Variables		Yes		Yes		Yes	
District Fixed Effects		Yes		No		No	
Community Fixed Effects		No		Yes		Yes	
Week Fixed Effects		Yes		Yes		Yes	

Notes: This table reports the coefficients of housing attributes of the STAR$(2,0)$ specification of (2.32) and (2.33). Other control variables include indicator variables for floors, building types, decoration condition, building structures, and elevator. Standard errors are reported in parentheses.

Table 2.9: Coefficients of Housing Attributes Obtained from $\texttt{STAR}(2,2)$ Model

		QMLE		AQSE		RAQSE	
Sample	Variables	Coefficients (1)	SE (2)	Coefficients (3)	SE (4)	Coefficients (5)	SE (6)
Mar-April	λ_{cmty}	-0.00465	(0.00104)	0.017244	(0.00161)	-0.00221	(0.00097)
	λ_{nbr}	0.002437	(0.00620)	0.124535	(0.02009)	0.013425	(0.02111)
	ρ_{cmty}	0.579368	(0.01105)	0.251791	(0.03149)	0.462609	(0.02270)
	ρ_{nbr}	0.269303	(0.00951)	0.144149	(0.03235)	0.476467	(0.02249)
	Living area (square meters)	0.015513	(0.00017)	0.014703	(0.00023)	0.015625	(0.00051)
	Living area^2	-3.2E-05	(5.5E-07)	-2.9E-05	(7.6E-07)	-3.2E-05	(2.3E-06)
	Age	-0.00440	(0.00027)	0.004748	(0.00033)	-0.00383	(0.00041)
	Number of bedrooms	0.038607	(0.00267)	0.044913	(0.00373)	0.037900	(0.00340)
	Number of living rooms	0.034234	(0.00294)	0.029209	(0.00418)	0.030376	(0.00315)
	Number of kitchens	0.137752	(0.01529)	0.156796	(0.02187)	0.140728	(0.05273)
	Number of bathrooms	-0.02883	(0.00403)	-0.02100	(0.00573)	-0.03055	(0.00652)
May-Jun	λ_{cmty}	0.001414	(0.00106)	0.019328	(0.00202)	0.000538	(0.00108)
	λ_{nbr}	0.147218	(0.00655)	0.138294	(0.02007)	-0.02137	(0.03076)
	ρ_{cmty}	0.540120	(0.01057)	0.256729	(0.03647)	0.415456	(0.02816)
	ρ_{nbr}	0.330225	(0.00962)	0.142117	(0.03071)	0.530048	(0.02934)
	Living area (square meters)	0.016588	(0.00016)	0.015270	(0.00023)	0.016256	(0.00051)
	Living area^2	-3.5E-05	(5.4E-07)	-2.9E-05	(8.3E-07)	-3.4E-05	(2.1E-06)
	Age	-0.00335	(0.00028)	0.006848	(0.00033)	-0.00308	(0.00047)
	Number of bedrooms	0.031392	(0.00270)	0.031848	(0.00378)	0.032329	(0.00395)
	Number of living rooms	0.032461	(0.00286)	0.029432	(0.00402)	0.028674	(0.00300)
	Number of kitchens	0.298551	(0.01386)	0.272176	(0.01930)	0.293797	(0.06571)
	Number of bathrooms	-0.01676	(0.00391)	-0.01515	(0.00546)	-0.01916	(0.00665)
Jul-Aug	λ_{cmty}	0.015776	(0.00111)	0.013718	(0.00189)	-0.00070	(0.00137)
	λ_{nbr}	0.213803	(0.00717)	0.136857	(0.02055)	-0.03741	(0.02579)
	ρ_{cmty}	0.521496	(0.01084)	0.253569	(0.03183)	0.397460	(0.02511)
	ρ_{nbr}	0.343574	(0.01008)	0.149156	(0.02845)	0.562784	(0.02613)
	Living area (square meters)	0.014861	(0.00017)	0.013595	(0.00022)	0.014340	(0.00054)
	Living area^2	-2.7E-05	(5.1E-07)	-2.3E-05	(6.9E-07)	-2.6E-05	(2.3E-06)
	Age	-0.00477	(0.00031)	0.006635	(0.00036)	-0.00418	(0.00041)
	Number of bedrooms	0.040316	(0.00295)	0.042903	(0.00396)	0.041173	(0.00410)
	Number of living rooms	0.043482	(0.00315)	0.040437	(0.00425)	0.039604	(0.00350)
	Number of kitchens	0.252999	(0.01511)	0.259863	(0.02020)	0.245021	(0.05318)
	Number of bathrooms	-0.02969	(0.00423)	-0.01847	(0.00584)	-0.02907	(0.00706)
Sep-Oct	λ_{cmty}	0.000994	(0.00119)	0.013345	(0.00171)	-0.00047	(0.00121)
	λ_{nbr}	-0.01639	(0.00791)	0.149845	(0.02202)	-0.05325	(0.03473)
	ρ_{cmty}	0.577041	(0.01116)	0.224911	(0.02871)	0.375436	(0.02804)
	ρ_{nbr}	0.279090	(0.01003)	0.159910	(0.02566)	0.574905	(0.02933)
	Living area (square meters)	0.013959	(0.00018)	0.012978	(0.00025)	0.013785	(0.00066)
	Living area^2	-2.5E-05	(5.5E-07)	-2.1E-05	(7.8E-07)	-2.5E-05	(2.7E-06)
	Age	-0.00454	(0.00031)	0.005885	(0.00038)	-0.00451	(0.00046)
	Number of bedrooms	0.050790	(0.00304)	0.055550	(0.00428)	0.049712	(0.00472)
	Number of living rooms	0.039543	(0.00327)	0.040360	(0.00472)	0.041181	(0.00420)
	Number of kitchens	0.134800	(0.01532)	0.118975	(0.02156)	0.131567	(0.04658)
	Number of bathrooms	-0.02855	(0.00456)	-0.02134	(0.00652)	-0.02912	(0.00712)
Nov-Dec	λ_{cmty}	0.000325	(0.00100)	0.027002	(0.00233)	0.002953	(0.00168)
	λ_{nbr}	0.100034	(0.00591)	0.136576	(0.01734)	-0.01241	(0.03120)
	ρ_{cmty}	0.551367	(0.00988)	0.290670	(0.03673)	0.465579	(0.02907)
	ρ_{nbr}	0.311276	(0.00873)	0.135637	(0.02822)	0.503385	(0.02959)
	Living area (square meters)	0.014518	(0.00013)	0.013742	(0.00019)	0.014269	(0.00055)
	Living area^2	-2.8E-05	(4.1E-07)	-2.5E-05	(6.0E-07)	-2.7E-05	(2.3E-06)
	Age	-0.00507	(0.00025)	0.005477	(0.00030)	-0.00423	(0.00041)
	Number of bedrooms	0.043774	(0.00230)	0.045463	(0.00316)	0.045426	(0.00346)
	Number of living rooms	0.039471	(0.00244)	0.034677	(0.00353)	0.035428	(0.00307)
	Number of kitchens	0.139307	(0.01162)	0.124396	(0.01615)	0.123489	(0.03095)
	Number of bathrooms	-0.02403	(0.00344)	-0.02360	(0.00480)	-0.02609	(0.00546)
Other Control Variables		Yes		Yes		Yes	
District Fixed Effects		Yes		No		No	
Community Fixed Effects		No		Yes		Yes	
Week Fixed Effects		Yes		Yes		Yes	

Notes: This table reports the coefficients of housing attributes using the $\texttt{STAR}(2,2)$ specification of (2.32) and (2.33). Other control variables include indicator variables for floors, building types, decoration condition, building structures, and elevator. Standard errors are reported in parentheses.

Table 2.10: Sample Size by Month

(1) Month	(2) Sample Size	(3) Percent
1	4,915	4.76%
2	2,764	2.68%
3	8,755	8.49%
4	9,735	9.44%
5	9,436	9.15%
6	9,712	9.42%
7	9,170	8.89%
8	8,929	8.66%
9	7,277	7.05%
10	8,462	8.20%
11	10,042	9.74%
12	13,952	13.53%
Total	103,149	100%

Appendix

Appendix A: Technical Assumptions

To discuss the asymptotic properties of the proposed QML, AQS and RAQS estimators of the $\mathtt{STAR}(p,q)$ model, it is convenient to denote the true value of a parameter by adding a subscript '$_0$'. A set of conditions under which our asymptotic results hold are listed below.

Assumption A1: *The true value δ_0 of the vector of spatial parameters is in the interior of a compact parameter set $\Delta = \Delta_\lambda \times \Delta_\rho$.*[21]

Assumption A2: $\{\varepsilon_{N,i}\}$ are iid$(0, \sigma_0^2)$, and $E|\varepsilon_{N,i}|^{4+\varepsilon} < \infty$ for some $\varepsilon > 0$.

Assumption A2*: $\varepsilon_N \sim (0, \sigma_0^2 H_N)$, *where* $H_N = \mathtt{diag}(h_{N,1}, \ldots, h_{N,N})$, $h_{N,i} > 0, \forall i$ *and* $\frac{1}{N}\sum_{i=1}^{N} h_{N,i} = 1$, *and* $E|\varepsilon_{N,i}|^{4+\eta} < c$ *for some* $\eta > 0$ *and constant c for all N and i.*[22]

Assumption A3: *C and T are fixed. The elements of $\mathbb{X}_N(\rho_0)$ are uniformly bounded for all N, $\mathbb{X}_N(\rho_0)$ has the full rank $k+C+T-1$, and $\lim_{n\to\infty} \frac{1}{N}\mathbb{X}_N'(\rho_0)\mathbb{X}_N(\rho_0)$ exists and is non-singular.*

Assumption A3*: *C and/or T grows with N. The elements of $\mathbb{X}_1(\rho_0)$ are uniformly bounded for all N, $\mathbb{X}_1(\rho_0)$ has the full rank k, and $\lim_{n\to\infty} \frac{1}{N}\mathbb{X}_N'(\rho_0)Q_2(\rho_0)\mathbb{X}_1(\rho_0)$ exists and is non-singular.*

Assumption A4: *The spatial weights matrices $W_{\ell j}$ and W_{ek} are uniformly bounded in both row and column sum norms and their diagonal elements are zero.*

Assumption A5: *The matrix $A_N(\lambda_0)$ and $B_B(\rho_0)$ are non-singular and $A_N^{-1}(\lambda_0)$ and $B_N^{-1}(\rho_0)$ are uniformly bounded in both row and column sum norms. Further,*

[21] For the log-likelihood function to be well defined, the parameter space Δ must be such that $A_n(\lambda)$ and $B_n(\rho)$ are non-singular $\forall \delta \in \Delta$. Lee and Liu (2010) show that since $\|\sum_{j=1}^{p}\lambda_j W_{\ell j}\| \le (\sum_{j=1}^{p}|\lambda_j|)\cdot \max_{j=1,\ldots,p}\|W_{\ell j}\|$, a viable parameter space for λ_j is such that $\sum_{j=1}^{p}|\lambda_j| < (\max_{j=1,\ldots,p}\|W_{\ell j}\|)^{-1}$, which simplifies to $\sum_{j=1}^{p}|\lambda_j| < 1$ when $W_{\ell j}$ are row-normalised. However, Elhorst et al. (2012) argue that this parametrisation is too restrictive and give an alternative procedure to determine exact boundaries which depends on the specification of $W_{\ell j}$. Similar arguments apply for the parameter space of $\{\rho_k\}$.

[22] For generality, we allow $h_{N,i}$ to depend on N for each i, which is sensible as heteroskedasticity may depend on the degree of spatial interaction, e.g., number of neighbours. This parametrisation is a non-parametric version of Breusch and Pagan (1979) and allows the estimation of the average scale parameter.

$A_N^{-1}(\lambda)$ *and* $B_N^{-1}(\rho)$ *are uniformly bounded in either row sum norms or column sum norm, uniformly in* $\lambda \in \Delta_\lambda$ and $\rho \in \Delta_\rho$.

The above are standard regularity conditions, extending directly those for the higher-order spatial autoregressive models of Liu and Yang (2017). What is left is the identification uniqueness conditions. The cases of fixed C and T fit into the framework of Liu and Yang (2017) and their Assumption 6 (for homoskedastic case) and Assumption 6* (for the heteroskedastic case) can be used. These conditions can be modified to suit our large C and/or T cases. The details are available from the authors upon request.

Appendix B: Some Technical Details

Lemmas B.1-B.5 below are the extended versions of the selected lemmas from Lee (2004), Kelejian and Prucha (2001) and Lin and Lee (2010).

Lemma B.1. *Let $\mathbb{X}_n(\rho)$ be defined in Sec. 3. Let $\mathbb{P}_N(\rho) = \mathbb{X}_N(\rho)[\mathbb{X}'_N(\rho)\mathbb{X}_N(\rho)]^{-1}\mathbb{X}'_N(\rho)$ and $\mathbb{Q}_N(\rho) = I_N - \mathbb{P}_N(\rho)$ be two projection matrices. Under Assumptions A1,A3 and A4, for each $\rho \in \Delta_\rho$, $\mathbb{P}_N(\rho)$ and $\mathbb{M}_N(\rho)$ are uniformly bounded in both row and column sum norms.*

Lemma B.2. *Let A_N and B_N be two $N \times N$ matrices, uniformly bounded in both row and column sum norms. Then for $\mathbb{Q}_N(\rho)$ defined in Lemma B.1, we have for each $\rho \in \Delta_\rho$,*

(i) $\mathtt{tr}(A_N^m) = O(n)$ *for* $m \geq 1$,

(ii) $\mathtt{tr}(A'_N A_N) = O(N)$,

(iii) $\mathtt{tr}((\mathbb{Q}_N(\rho)A_N)^m) = \mathtt{tr}(A_N^m) + O(1)$ *for* $m \geq 1$,

(iv) $\mathtt{tr}((A'_N \mathbb{Q}_N(\rho)A_N)^m) = \mathtt{tr}((A'_N A_N)^m) + O(1)$ *for* $m \geq 1$,

(v) $A_N B_N$ *is uniformly bounded in both row and column sum norms,*

(vi) $\mathtt{tr}(A_N B_N) = \mathtt{tr}(B_N A_N) = O(N)$ *uniformly.*

Lemma B.3. *(Moments and Limiting Distribution of Quadratic Forms): For a given process of innovations $\{\varepsilon_{N,i}\}$, assume $\varepsilon_{N,i} \sim inid(0, \sigma_0^2 h_{N,i})$, where $h_{N,i} > 0$ and $\sum_{i=1}^N h_{N,i} = N$. Let $H_N = diag(h_{N,1}, \ldots, h_{N,N})$, D_{rN} be $N \times N$ matrices of elements $d_{rN,ij}$, and c_{rN} $N \times 1$ vectors of elements $c_{rN,i}$. For $Q_{rN} = \varepsilon'_N D_{rN} \varepsilon_N + c'_{rN}\varepsilon_N$ where $r = 1, \ldots, R$, we have,*

(i) $\mathrm{E}(Q_{rN}) = \sigma_0^2 \mathtt{tr}(H_N D_{rN})$,

$$(ii)\ \mathrm{Var}(Q_{rN}) = \sigma_0^4 \mathtt{tr}[H_N D_{rN}(H_N D_{rN} + H_N D'_{rN})] + \sigma_0^2 c'_{rN} H_N c_{rN} + \textstyle\sum_{i=1}^n (\sigma_{N,i}^4 d_{rN,ii}^2 \kappa_{N,i} + 2\sigma_{N,i}^3 d_{rN,ii} c_{N,i} \gamma_{N,i})$$

$$(iii)\ \mathrm{Cov}(Q_{rN}, Q_{sN}) = \sigma_0^4 \mathrm{tr}(H_N D_{rN}(H_N D_{sN} + H_N D'_{sN})) + \sigma_0^2 c'_{rN} H_N c_{sN} + \textstyle\sum_{i=1}^N [\sigma_{N,i}^4 d_{rN,ii} d_{sN,ii} \kappa_{N,i} + \sigma_{N,i}^3 (d_{rN,ii} c_{sN,i} + d_{sN,ii} c_{rN,i}) \gamma_{N,i}],$$

where $\sigma_{N,i} = \sigma_0 \sqrt{h_{N,i}}$, $\gamma_{N,i}$ = skewness of $\varepsilon_{N,i}$, and $\kappa_{N,i}$ = excess kurtosis of $\varepsilon_{N,i}$.

Lemma B.4. *Under Lemma B.3, if further D_{rN} is uniformly bounded in either* *row or column sum norm, then, we have, for* $r = 1, \ldots, m$,

(i) $\mathrm{E}(Q_{rN}) = O(N)$, (ii) $\mathrm{Var}(Q_{rN}) = O(N)$,

(iii) $Q_{rN} = O_p(N)$, (iv) $\frac{1}{n}[Q_{rN} - \mathrm{E}(Q_{rN})] = O_p(N^{-\frac{1}{2}})$.

Lemma B.5. *Under Lemma B.3, let* $\mathbf{Q}_N = (Q_{1N}, \ldots, Q_{mN})'$ *and* $\mathbf{\Sigma}_N = \mathrm{Var}(\mathbf{Q}_n)$. *Assume* $\mathbf{\Sigma}_N^{-1}$ *exists and denote its square-root matrix by* $\mathbf{\Sigma}_N^{-1/2}$. *If further* D_{rN} *is uniformly bounded in both row and column sum norms, and Assumptions 2* holds for* $\{\varepsilon_{Ni}\}$, *then we have,*

(i) $\frac{Q_{rN} - \mathrm{E}(Q_{rn})}{\sqrt{\mathrm{Var}(Q_{rN})}} \xrightarrow{D} \mathcal{N}(0,1)$, $r = 1, \ldots, m$, *and* (ii) $\mathbf{\Sigma}_N^{-1/2}(\mathbf{Q}_N - \mathrm{E}(\mathbf{Q}_N)) \xrightarrow{D} \mathcal{N}(0, I_N)$.

Chapter 3

Urbanization, Feminization, and the Marriage Market

3.1 Introduction

Starting from Marshall (1890), researchers have put numerous efforts into the area of agglomeration economies, whereby the cluster of firms and workers generates positive externalities to other firms and workers (Duranton et al. 2015). It is very natural that majority of the research focusing on the micro-foundation or the benefits of agglomeration economies is more about the gain in productivity and the subsequent labour market outcomes (Duranton and Puga 2004). The incentives arising from the matching outcomes in the marriage market have attracted less attention, although it can be an equally important channel for the process of urbanization. In this paper, we study this under explored migration incentives via utilizing the quasi-experimental settings created by the setup of Special Economics Zones (SEZs) in China, and China's accession to the World Trade Organization (WTO). We combined information of the quasi-experiments wih individual level data on migration and marital status obtained from the Population Census in China to achieve the identification.

Understanding the migration incentives arising from the matching outcomes in the marriage market, especially the heterogeneous responses between females and males to the incentives is crucial. The phenomenon of "missing women" resulting from sex-selective abortion, neglect or infanticide in China and other developing counties has been well documented (Qian 2008; Almond et al. 2019; Duflo 2012; Lin et al. 2014; Barcellos et al. 2014). The gender imbalance in rural ares is particularly severe. The imbalance would be further strengthened if females in the rural areas are more likely to migrate to urban areas than males, which will eventually affect the marriage market outcomes. As the marriage market outcomes have non-negligible influences on fertility, inequality and income redistribution (Fernández and Rogerson 2001), it justifies the importance to figure out the explicit migration incentives for males and females. Moreover, studying this question also helps to better understand the insights of the marital matching. On the one hand, the positive marital assortative matching based on income or education is documented to raise inequality (Greenwood et al. 2014; Hryshko et al. 2017). On the other hand, females are usually easier to "marry up" to persons whose socio-economic origins are higher than themselves, compared with their males counterparts (Edlund 2005; Charles and Luoh 2010; Qian 2008). Thus, if the setup of SEZs increase the matching outcomes of the rural-urban female migrants, it may function as a channel that reduces the inequality.

We examine the causal impacts of the changes of the incentives created by the setup of SEZs and China's WTO accession on individuals' migration patterns, especially the heterogeneous effects across different genders. We investigate this question by utilizing the quasi-experiment of the establishment of the Special Economic Zones in China since 1980s that exogenously changed the incentives for migration, attributing to the increased work opportunities and probability to meet up with other people. The setup of SEZs can be viewed as a pull force that attract migrant workers to the places where SEZs were granted. To consolidate our findings, we also exploit another exogenous variations in tariff reductions that caused by China's accession to the WTO. We construct a variable using the tariff reductions together with pre-WTO-accession migration flows to measure the shock from China's accession to the WTO as a push force, which makes people have larger tendency to leave their hometown.

We conduct the analysis in three steps. We first examine the impacts of the establishment of SEZs on the number of in-bound migrants at the county level. We then further study the heterogeneous effects on migration patterns for males and females by utilizing the individual level Census data. Finally, we examine the impacts on the marriage market together with the heterogeneity across genders. Our findings provide evidence for the existence of the incentives arising from the marriage market which has attracted less attention in the literature.

We obtain the following results. First, we find that the setup of SEZs increase the population of the counties where SEZs are located. On average, there exists a 3.1% increase in total population. In particular, the increase in population can be largely attributed to the increase in in-bound migrants. The number of in-bound migrants increase by 17%. Moreover, the impacts are heterogeneous on male and female migrants. We find a sharp and significant increase (up to 40%) of young single female migrants, while the magnitude for young single males is 14%. These results are obtained from using the classic staggered difference-in-differences model with two-way fixed effects. The results remain when we adopt the methods proposed in Callaway and Sant'Anna (2021) and Sant'Anna and Zhao (2020). Second, empowered by the individual level Population Census data, we verify that there is a unique feminization phenomenon during this process. Among all single migrants, the proportion of single female migrants increases significantly by 3% due to the SEZ establishment. We further find that the effects are only for those who are single and less educated, but not for migrants who are married or with higher educational levels. This finding is robust when we use China's accession to the WTO as the exogenous shock. Third, we find that there are indeed positive and significant effects on the marriage market outcomes, where female migrants gain advantages in the marriage market. The setup of SEZs make it easier for them to get married than their male counterparts.

This paper contributes to the literature on three grounds. First, this paper joins the literature on the benefits of agglomeration economies. Starting from Marshall (1890), researchers have reached the consensus over time that cities offer many benefits for workers, firms, and consumers including higher efficiency and better opportunities. Cities are usually associated with higher productivity, wages, and better amenities (Rosenthal and Strange 2004; Combes and Gobillon 2015; Diamond 2016; Couture and Handbury 2017). These benefits are created from the sharing of goods, people and ideas (Duranton and Puga 2004; Holmes 1999; Glaeser and Mare 2001; Moretti 2004; Ellison et al. 2010). At the same time, these benefits trigger urbanization and provide incentives for people to migrant. In our paper, we document the benefits and incentives arising from the better matching outcomes in the marriage market. We show that it is also an important factor that triggers urbanization.

Second, we contribute to the literature on the matching patterns in the marriage market by documenting a unique feminization phenomenon. As stated in Browning et al. (2014), "Marriages are not formed randomly. Rather, individuals sort themselves into marriage based on the attributes of both partners because interactions in individual attributes generate mutual gains from marriage". Many studies have documented the phenomenon of assortative matching on different characteristics including income, wages, education, personality traits or even body shapes (Becker 1991; Grossbard-Schectman and Grossbard-Shechtman 1993; Browning et al. 2014; Choo and Siow 2006; Pencavel 1998; Chiappori et al. 2012; Dupuy and Galichon 2014). The positive marital assortative matching on income and education typically raises inequality (Greenwood et al. 2014; Hryshko et al. 2017). However, on the other hand, the fecundity of young women makes them relatively scarce in the marriage market. As a result, they can potentially enjoy the scarcity rents to "marry up" with persons whose socio-economic origins are higher than them (Edlund 2005; Hamilton and Siow 2007). Our findings of the unique feminization phenomenon that young female migrants have advantages in obtaining better outcomes in the marriage market is in line with this theory.

Third, we contribute to the literature on place-based policies. It is inconclusive whether a place-based policy is effective or not (Neumark and Simpson 2015; Glaeser and Gottlieb 2008; Elvery 2009; Ham et al. 2011; Lynch and Zax 2011; Neumark and Kolko 2010). In the context of place-based policies in developing countries such as China, many researchers find that the experiments of setting up

SEZs since the late 1970s greatly increased the FDI, together with the local labor market outcomes such as output, employment, wages, and productivity (Wang 2013; Alder et al. 2016; Lu et al. 2019; Xu 2011). Our paper investigate the impacts of the setup of SEZs from a different angle. We study how the changes in incentives resulted from SEZ establishment triggered urbanization, and the heterogeneous effects across genders.

The remainder of the paper is organized as follows. Section 2 discusses the historical background of the SEZs and China's accession to the WTO. Section 3 provides the empirical framework for the analysis. Section 4 describes the data and variables. It also provides some relevant summary statistics. Section 5 reports the impacts of SEZs and WTO accession on population and migration patterns. Section 6 concludes.

3.2 Background

In this section, we briefly introduce the historical background of the two quasi-experimental settings we utilize in this paper – the establishment of the Special Economic Zones in China since the late 1970s and China's accession to the WTO in 2001.

3.2.1 The Establishment of Special Economic Zones

The establishment of Special Economic Zones was in a context of China's "reform and opening up" since 1978. The central government of China started the program of economic reforms in December 1978 to revitalize the disrupted economy after the Cultural Revolution, which gradually transformed the Soviet-type centrally planned economy into a market economy (Shirk et al. 1993).

The experiment of setting up SEZs is one of the most important components in the economic reforms.[1] The first four SEZs, located in two coastal provinces (Guangdong and Fujian), were formally approved by the central government to be established in summer 1979 (Crane 1990).

[1]There are other well-known parts of the reform such as the implementation of the Household Responsibility System (HRS) (Xu 2011).

As the experiment of setting up SEZs turned out to be successful, the central government further granted SEZs to fourteen coastal cities and the South China island of Hainan in 1984, and to the Pearl River Delta region, Yangtze River Delta region, and southern Fujian in 1985 (Shirk et al. 1993). Later, SEZs were granted to other regions in China on a regular basis, and were expanded from coastal areas to inland areas. Figure 3.5 shows the trajectory of establishing the Special Economic Zones. Figure 3.6 displays the numbers of SEZs established in each year from 1984 to 2017. There are significantly more SEZs established since 1990. The number of new SEZ setups fluctuated, and the number peaked in 1992-1993, 2003-2003, 2006, and 2010-2012. By 2017, 1915 counties in our dataset has at least one SEZ, which takes up around 67% of all the counties in China.

As the direct purpose of setting up SEZs is to attract foreign direct investment (FDI) and expand international trade (Shirk et al. 1993; Cai et al. 2008; Xu 2011), privileges such as concessionary customs and tax treatment were granted to these SEZs by the government. The first detailed legal rules on SEZs named “the Regulation for Guangdong SEZs" were issued in 1980.[2] Details about the preferential policies for foreign investors were listed in the third chapter of the regulation, including (1) preferential policies about land-use terms and fees; (2) import duty exemptions of raw materials, intermediates, and capital equipment for firms in the SEZs; (3) preferential corporate income tax at 15%, while the rates for firms outside of the SEZs were 33%; etc.[3]

As a result, these SEZs greatly increased the FDI, together with the local labor market outcomes such as output, employment, wages, and productivity (Wang 2013; Alder et al. 2016; Lu et al. 2019; Xu 2011). Moreover, it is equally important that the agglomeration economy arising from SEZs may also create better matching outcomes. It can potentially provide incentives to workers just like the labor market.

[2]The regulation is available at https://zh.wikisource.org/zh-hant/%E5%B9%BF%E4%B8%9C%E7%9C%81%E7%BB %8F%E6%B5%8E%E7%89%B9 %E5%8C%BA%E6%9D%A1%E4%BE%8B.

[3]Wang (2013), Lu et al. (2019) and Alder et al. (2016) have detailed discussions about the preferential policies for SEZs.

3.2.2 The Accession to the World Trade Organization

The accession to the WTO in 2001 is another milestone in the process of China's "reform and opening up". Related to what has been discussed in the last session, Figure 3.6 shows that many SEZs were established from 1980 to 2000. Concurrently, to meet the conditions of the WTO accession, the preferential policies gradually reduced tariff and non-tariff barriers unilaterally, which laid the groundwork for the entry into the WTO (Brandt et al. 2017; Lu and Yu 2015).

China's accession to the WTO further reduced the tariffs and non-tariff barriers. The unweighted average tariffs dropped by 20% from 15.3% in 2001 to 12.3% in 2004 (Lu and Yu 2015). More specifically, China's import tariffs on intermediate inputs fell sharply from around 14% in 2001 to 8% in 2004 (Ma et al. 2019). China also started to remove the non-tariff barriers by signing the WTO accession protocol, such as quotas, licenses and tendering requirements, by the end of 2004 (Imbruno 2016). It greatly reduced the trade policy uncertainties besides tariffs.

The reduced tariff and trade policy uncertainties had greatly promoted the economics growth in China. There were rapid increases in both exports and imports (Imbruno 2016; Yu et al. 2020). Total exports increased seven-fold from 2001 to 2016. The productivity growth was rapid since then (Brandt et al. 2012), Closely related to the productivity growth, firms' patent applications increased as well (Liu and Ma 2020).

Moreover, it also had significantly positive impacts on the outcomes in the labor market. The WTO accession created a vast amount of jobs, especially in the manufacturing industry, and in coastal and urban areas. As a result, a lot of workers move from rural, less developed inland areas to urban and coastal areas. The magnitude of the internal migration in China was unprecedented. Figure 3.1 shows the number of migrants reported in different census waves, where the total number of migrants almost doubled from around half a million in 2000 to more than one million in 2005.[4] Similar to the impacts arising from SEZs, the accession to the WTO can potentially create agglomeration economy, which provides another source of

[4]The migrants are defined as inter-county movers who are above 16, and whose residential address is different from the Hukou address.

incentives to the workers in the marriage market, and the responses to the incentives may be heterogeneous across genders.

3.3 Empirical Framework

We are interested in two key questions. First, do people respond to the changes of the incentives in the labor and the marriage markets? Second, are the effects the same for males and females? Our key identification strategy is to exploit the setup of SEZs and China's accession to the WTO as a pull force and push force that exogenously trigger urbanization across locations. These policy changes and events provide us with exogenous variation of the incentives that affect people's migration decisions. It enables us to investigate the effects on the migrants patterns, and check if there exists heterogeneity between males and females.

Next, we list the specifications we use to analyze the effects of SEZs and China's accession to the WTO. For the effects of SEZs, we adopt a staggered difference-in-differences model with two-way fixed effects. The base line specification at the county level is given by

$$y_{c,t} = \beta \mathbb{1}(\mathrm{SEZ}_{c,t}) + \alpha_c + \alpha_{p,t} + \varepsilon_{c,t}, \tag{3.1}$$

where $y_{c,t}$ is the outcome variable, including the logarithm of number of population and migrants at the county level. $\mathbb{1}(\mathrm{SEZ}_{c,t})$ captures the treatment status, where it is equal to 1 if county c has at least one SEZ at year t. β is our parameter of interest, representing the effects of SEZ setup. α_c and $\alpha_{p,t}$ are the county and province-by-year fixed effects, respectively.

To investigate whether females respond more to the shocks, we further conduct the analysis at the individual level, where the specification is as follows.

$$\mathbb{1}(\mathrm{Female})_{i,c,t} = \lambda \mathbb{1}(\mathrm{SEZ}_{c,t}) + \alpha_c + \alpha_{age} + \alpha_{p,t} + \varepsilon_{i,c,t} \tag{3.2}$$

where $\mathbb{1}(\mathrm{Female})_{i,c,t}$ is a gender indicator, which equals to 1 if an individual is a female. λ captures the heterogeneous effects of SEZ establishment across genders. If females and males share the same level of responses to the shocks caused by SEZs, λ will be 0. α_{age} are the age fixed effects. $\mathbb{1}(\mathrm{SEZ}_{c,t})$, α_c, and $\alpha_{p,t}$ are the same as in equation (3.1).

We also utilize the exogenous shock created by China's accession to WTO as a push force to reassure the findings in Equation 3.2. The regression is specified as follows.

$$\mathbb{1}(\text{Female})_{i,c,t} = \lambda \times WTOshock_{cy} \times Post_t + \alpha_c + \alpha_{age} + \alpha_{p,t} + \varepsilon_{i,c,t}, \qquad (3.3)$$

where $\mathbb{1}(\text{Female})_{i,c,t}$ is a gender indicator as in Equation 3.2. $WTOshock_{cy}$ is a proxy for treatment intensity, which is to measure the exposure to the shock. Note that $WTOshock_{cy}$ is constructed at the prefecture-city level (the subscript of $WTOshock_{cy}$ represents "city"). We will elaborate the way we construct this variable in the next section. $Post_t$ is a dummy variable that equals 1 if the year t is 2001 or after. α_c, and $\alpha_{p,t}$ are the same as before.

Finally, to examine the impacts on the marriage market, we adopt the following regression specification:

$$\mathbb{1}(\text{Married})_{i,c,t} = \lambda \mathbb{1}(\text{SEZ}_{c,t}) + \gamma \mathbb{1}(\text{SEZ}_{c,t}) \times \mathbb{1}(\text{Female}_i) + \alpha_c + \alpha_{age} + \alpha_{p,t} + \varepsilon_{i,c,t}. \qquad (3.4)$$

The parameter γ is the parameter of interest, which represents the difference between male and female migrants in the probability of getting married. Besides $\mathbb{1}(\text{Married})_{i,c,t}$, we also investigate other outcome variables such as the dummy variable for marrying with a local resident, as a proxy for the marriage market outcomes.

3.4 Data, Variables and Summary Statistics

3.4.1 Data

We use three primary datasets. We first collect the data of the SEZ establishments from the National Development and Reform Commission of China.[5] It con-

[5]There are two versions of the lists of SEZs published by the National Development and Reform Commission of China. The first version was published in 2006, which covered SEZs established by that year. There was an updated version later in 2018. In this paper, we combine the two versions to have a full list of SEZs together with their setup year from 1984 to 2017. There is also information about the size, level and main industries in a SEZ. The lists can be found at http://www.gov.cn/zhengce/zhengceku/2018-12/31/5434045/files/6eea5e4b78a645c1a27c231b152792ef.pdf, and https://www.ndrc.gov.cn/xxgk/zcfb/gg/200704/W020190905487497735524.pdf.

tains the names of the SEZs together with the year of establishments. To match a SEZ with the county where it is located in, we crape the data using the name of a SEZ or the name of a SEZ's management committee from Baidu Maps (a main mapping service application in China).

The second dataset we use is the National Population Census in China. There are six waves of the National Population Census (1982, 1990, 2000, 2005 and 2010) covered in this paper. We obtain individuals' gender, age, marital status, occupation, and industry from the Census. It also provides us with information about the population if we aggregate the individual data. Moreover, we are able to find information about people's residential status from the 2000 Census onward, through which we can define the internal migrants in China and investigate the migration patterns.

The third dataset is the exposure to trade shocks caused by China's accession to the WTO. Following Yu et al. (2020), we combine three source to construct the metrics for measuring the exposure to trade shock, which are the 2002-2013 World Bank Trade Analysis and Information System (TRAINS) dataset, the Annual Survey of Industrial Firms (ASIF), and the 2000 Population Census.

3.4.2 Variables and Summary Statistics

In this section, we describe how we prepare our data for the empirical analysis and present the basic summary statistics.

Figure 3.2 plots the the sex ratio of migrants and local residents, which is defined as number of males divided by number of females. It is worth noting that the sex ratio of migrants plunges for young migrants. However, we do not observe such pattern for local residents. Moreover, we take a closer investigation on the numbers of female and male migrants, which are shown in Figure 3.3. There are more female migrants than male migrants if we compare migrants aged between 14 and 20. This pattern flips for older migrants. Figure 3.4 is the self-reported marriage rate by age of female and male migrants, showing that the gap of married rates between female and male migrants peak at around 25 years old. These summary statistics indicate that there may exist different incentives that make the migration decisions differ across genders.

For analyzing the impacts of SEZs on migration, we create a list of SEZs together with the setup year, and match them with the corresponding counties. Table 3.1 reports the number of total population, migrants and and sex ratios across years in the counties that ever had one SEZ established. It shows that there was an increase in the number of total population and migrants. At the same time, sex ratios in these counties decreased.

For analyzing the impacts of China's accession to the WTO, we construct a prefecture-city level[6] variable to measure cities' exposure to the shocks caused by the WTO accession. This variable is designed to be a proxy representing the extend to which the accession to the WTO provides a push force that encourages people to move to other cities to work. It is constructed in three steps. First, we calculate the prefecture-city-year specific tariff reduction over 2003-2013. It measures how attractive a city is to workers. Greater tariff reduction are usually associated with more job opportunities, as the tariff reduction is a direct measure of the trade shock. Second, for each prefecture-city-year, we calculate the sum of the tariff reduction of all other cities that have migration flow between the two cities, weighted by the share of the migration flow. The migrantion flow data are from the 2000 Population Census, which is the pre-WTO-accession period. Then we sum them up across years to obtain our measure of WTO shocks. Equation (3.5) shows how we construct the metric, where m_i is city i's total number of outbound migrants, and m_{ij} is the number of city i's outbound migrants that work in city j.[7]

$$\begin{aligned} WTOShock_{it} &= \sum_{j \neq i} tariff_{jt} \times \frac{m_{ij}}{m_i}, \\ WTOShock_i &= \sum_{t=2003}^{2013} WTOShock_{it}. \end{aligned} \tag{3.5}$$

[6] The counties (or county-level cities) are sub-units of prefecture-level cities in China's administrative hierarchy (Qin and Zhang 2014).

[7] See Yu et al. (2020) for a more detailed discussion of constructing the exposure to the WTO accession.

3.5 Results

3.5.1 The Impacts of SEZs on Population

We first examine the impacts of SEZs on population and number of migrants at the aggregate level (county level). Table 3.2 reports the standard two-way fixed effects (TWFE) estimation results of the effects of SEZs on population and migrants from estimating Equation (3.1). Column (1) of Table 3.2 suggests that on average, counties experienced a 3.1% increase in the total population.

To take a closer investigation on the components of the increase in population, we further restrict our samples to (all) migrants, female migrants, and male migrants, respectively. It enables us to compare the magnitudes of the effects of the SEZs across different groups of people, and figure out the potential key drivers of the population growth illustrated in column (1). Column (2)-(4) report the effects of SEZs on number of young migrants. We choose the dependent variables to be young migrants aged between 16-25, as they are most likely to be facing marital decisions, and are probably the most responsive to the incentives in the marriage market, as shown in Figure 3.4.[8] Column (2) suggests that there is a 17% increase in number of inbound migrants after the SEZ establishment. Column (3) and (4) compares the magnitudes of the effects for single female and male migrants. We find a 40% increase in single female migrants, while the effect for single male migrants are only 14% and insignificant.

Figure 3.7 plots the dynamic effects of SEZs on the number of single female and male migrants. There are two main findings. First, we do not observe evidence for pre-treatment non-parallel trends. Second, not surprisingly, the magnitudes of the effects for single female migrants are larger than that of male migrants.

Panel B.1 and Panel B.2 of Table 3.2 report the results using the method proposed by Callaway and Sant'Anna (2021) and Sant'Anna and Zhao (2020). Their estimation procedures for the staggered DiD setups are designed to circumvent the

[8]The legal marriageable age in China is 22 for males and 20 for females. There exists married people younger than the marriageable age in Figure 3.4, because the marital status in the census is self-reported. Early marriage is more prevalent in developing countries, especially for less developed countries and rural areas (Singh and Samara 1996).

"negative weighting problem" discussed in Goodman-Bacon (2021).[9] Thus, we are able to obtain more credible results of the effects on migrants. Panel B1 shows the results of using "not-yet-treated" groups as the control groups. The magnitudes of the effects on total population remain almost the same as the TWFE results. Column (2) shows that the magnitude of the effect of SEZs on all migrants are much larger compared with the TWFE results. There is a 48% increase in the number of migrants for the treatment group. The results of single female migrants and single male migrants in column (3) and (4) also show the same pattern. It is worth noting that the effects on single male migrants become significant, but the magnitude is still smaller than that of female migrants. Panel B.2 is the estimation results using "never-treated" groups as the control groups. The results are similar as in Panel B.1. Figure 3.8 plots the corresponding dynamic effects of the SEZs on population and migrants. Panel A shows that the effects on population is not very salient in the next few years after the establishment of SEZs. The effects for migrants, single female migrants and male migrants are shown in Panel B, C and D. There are clear patterns that SEZs increase number of inbound migrants, and the magnitudes grow with time.

3.5.2 The Impacts of SEZs on Gender Balance

There are suggestive results in Table 3.2 showing that single female migrants are more responsive to the incentives of SEZs. In this section, we further investigate the differential effects of SEZs on single male and female migrants by estimating a model at individual level. Specifically, we use the dummy variable for female as the dependent variable. It allow us to estimate a DiD model at the individual level as in Equation (3.2) to investigate the changes in gender balance before and after a SEZ establishment. Moreover, we present the estimation results of individuals with different education level and marital status to have a clearer understanding of the potential channels through which the SEZs affect young people's migration decisions. It also enables us to verify if the marriage market is playing a role.

[9]The idea is to choose proper control groups that are free from the "negative weighting problems", and identify the disaggregated "group-time average treatment effects".

To check the changes in gender balance before and after a SEZ establishment, we first estimate a staggered DiD model at individual level as in Equation (3.2). The results are presented in Table 3.3. Column (1) shows that the effects of SEZs on the female indicator are positive and significant. It means conditional on young single migrants, the percentage of females tend to increase after the SEZ establishment.[10] However, we do not obtain the same results for married migrants, as shown in column (4), which indicates that the pattern in column (1) can potentially be attributed to the incentive in the marriage market. As a result, we observe that single migrants are more responsive.

Next, we further divide our sample in column (1) into two groups – migrants with low education and high education.[11] The results are presented in column (2) and (3). It shows that although the magnitude of the effects are almost the same for migrants with low education and high education, the results are more significant for the former group. This finding is also in line with the prediction in Edlund (2005) and Hamilton and Siow (2007). We conduct additional robustness checks by adding migrants' working industry and occupation fixed effects. Our finding remains, and the results are presented in Table 3.4.

3.5.3 The Impacts of China's Accession to the WTO

In Section 3.5.1 and 3.5.2, we exploit the setup of special economic zones as a pull force that triggers the urbanization process. We observe that females, especially young single females are more responsive than their male counterparts. We now present the evidence of the differential responses to the incentives across genders from another perspective. In this section, we utilize China's accession to the WTO as a push force to further consolidate our findings.

Again, we examine the impacts on gender balance by using the female dummy as our outcome variable. Our analysis in this section differs from Section 3.5.2 in two respects. First, the shock induced by China's accession to the WTO are used

[10] We focus on migrant workers who are from rural areas (with a rural Hukou).

[11] Migrants with low education are defined as people who have a junior high school education or below, while migrants with high education are defined as people who have a senior high school education or above.

as a push force that triggers migrants to move to other cities. On the contrary, the establishment of SEZs are a pull force that attract migrants. As a results, we expect a larger decrease in population after the WTO accession for places that received larger shocks, while we observe a larger increase after the shock of SEZs. Second, the shocks induced by China's accession to the WTO in this section are measured at the prefecture-city level due to the issue of data availability, while county level shocks are used in Section 3.5.1.

We adopt a similar regression specification as in Section 3.5.2 from estimating Equation (3.3). Column (1) in Table 3.5 shows that the effects of the WTO accession on the female indicator are positive and significant, which indicates single females are more responsive than their male counterparts. On the other side, there are no significant effects for married migrants as shown in column (4). If we further decompose the sample used in column (1) into single migrants with different educational levels, the results are reported in columns (2) and (3). It shows that low educated single females are affected more than high educated ones. Lastly, we further include industry and occupation fixed effects into the model. The results are presented in Table 3.6. The findings remain the same.

3.5.4 The Impacts on the Marriage Market Outcomes

We have documented the phenomenon that the females are more responsive to the incentives created by the SEZs or the WTO accession. We argue that the reason can be that migrants move not only because of the incentives in the labor market, but also the incentives in the marriage. Moreover, female migrants have advantages over male migrants as predicted and documented by Edlund (2005), Hamilton and Siow (2007), and Qian and Qian (2017). In this section, we examine the impacts of the SEZ establishment on the marriage market outcomes to check if the empirical findings support the hypothesis that females are more responsive because they have advantages in the marriage market.

Table 3.7 presents the results from estimating Equation 3.4. The dependent variable for column (1) and (2) is a dummy variable indicating the marital status of an individual. The parameter of interest is γ in Equation (3.4), which is the coefficient of the triple difference-in-differences estimator that measures the heterogeneous effects of SEZs on the probability of being married. It shows that the SEZ establishments are indeed having an positive effect for females, and the effect is significantly larger than male migrants. Column (3) and (4) reports the results of using the dummy variable of marrying to a local resident as the dependent variable. We also find that it is easier for female migrants to be married and/or marry to local residents than males.[12] The agglomeration economy not only provides incentives in the marriage market to female migrants that make them move to places with SEZs, but also improved the matching results.

3.6 Conclusion

The incentives for urbanization and migration arising in productivity gains and the subsequent labour market outcomes improvements has been studied intensively in the literature. However, there is less attention on understanding the incentives in the marriage market. In this paper, we study this under explored migration incentives via utilizing the quasi-experimental settings created by the setup of Special Economics Zones in China, and China's accession to the World Trade Organization. We combine them with individual level data on migration and marital status obtained from the Population Census in China to identify the differential impacts between males and females.

We find that the setup of SEZs increases the population of the counties where SEZs are located. In particular, the increase in population can be largely attributed to the increase in in-bound migrants. Moreover, the impacts are heterogeneous on male and female migrants. We find larger increase of young single female migrants than males. Also, empowered by the individual level Population Census data, we verify that there is a unique feminization phenomenon during this process. Among all single migrants, the proportion of single female migrants increases significantly due to the SEZ establishment. We further find that the effects are only for those who are single and less educated, but not for migrants who are married or with higher

[12] The coefficient γ in the column (4) is marginally significant, where the p-value is 0.11.

educational level. This finding is robust when we use China's accession to the WTO as the exogenous shock. Lastly, we find that there are indeed positive and significant effects on the marriage market outcomes, where female migrants gain advantages in the marriage market. The setup of SEZs make it easier for them to get married than their male counterparts. Taken together, these findings point to the existence of the incentives for urbanization and migration arising from the potentially improved outcomes in the marriage market. Specifically, the incentives are larger for females compared with males.

While our findings highlights the role of incentives arsing from the marriage market. Future work would benefit from carrying out a welfare analysis for different groups of people who experienced the shocks of the SEZ establishment.

References

Alder, S., Shao, L., and Zilibotti, F. (2016). Economic Reforms and Industrial Policy in a Panel of Chinese Cities. *Journal of Economic Growth*, 21(4):305–349.

Almond, D., Li, H., and Zhang, S. (2019). Land Reform and Sex Selection in China. *Journal of Political Economy*, 127(2):560–585.

Barcellos, S. H., Carvalho, L. S., and Lleras-Muney, A. (2014). Child Gender and Parental Investments in India: Are Boys and Girls Treated Differently? . *American Economic Journal: Applied Economics*, 6(1):157–89.

Becker, G. S. (1991). *A Treatise on the Family: Enlarged Edition*. Harvard university press.

Brandt, L., Van Biesebroeck, J., Wang, L., and Zhang, Y. (2017). WTO Accession and Performance of Chinese Manufacturing Firms. *American Economic Review*, 107(9):2784–2820.

Brandt, L., Van Biesebroeck, J., and Zhang, Y. (2012). Creative Accounting or Creative Destruction? Firm-level Productivity Growth in Chinese Manufacturing. *Journal of Development Economics*, 97(2):339–351.

Browning, M., Chiappori, P.-A., and Weiss, Y. (2014). *Economics of the Family*. Cambridge University Press.

Cai, D. et al. (2008). *Xi Zhongxun Governing Guangdong (Xi Zhongxun Zhuzheng Guangdong)*. Chinese Communist Party History Press, Beijing.

Callaway, B. and Sant'Anna, P. H. (2021). Difference-in-differences with Multiple Time Periods. *Journal of Econometrics*, 225(2):200–230.

Charles, K. K. and Luoh, M. C. (2010). Male Incarceration, the Marriage Market, and Female Outcomes. *The Review of Economics and Statistics*, 92(3):614–627.

Chiappori, P.-A., Oreffice, S., and Quintana-Domeque, C. (2012). Fatter Attraction: Anthropometric and Socioeconomic Matching on the Marriage Market. *Journal of Political Economy*, 120(4):659–695.

Choo, E. and Siow, A. (2006). Who Marries Whom and Why. *Journal of Political Economy*, 114(1):175–201.

Combes, P.-P. and Gobillon, L. (2015). The Empirics of Agglomeration Economies. In *Handbook of Regional and Urban Economics*, pages 247–348. Elsevier.

Couture, V. and Handbury, J. (2017). Urban Revival in America, 2000 to 2010. Technical report, National Bureau of Economic Research.

Crane, G. T. (1990). *The Political Economy of China's Economic Zones*. Routledge.

Diamond, R. (2016). The Determinants and Welfare Implications of Us Workers' Diverging Location Choices by Skill: 1980-2000. *American Economic Review*, 106(3):479–524.

Duflo, E. (2012). Women Empowerment and Economic Development. *Journal of Economic Literature*, 50(4):1051–79.

Dupuy, A. and Galichon, A. (2014). Personality Traits and the Marriage Market. *Journal of Political Economy*, 122(6):1271–1319.

Duranton, G., Henderson, V., and Strange, W. (2015). *Handbook of Regional and Urban Economics*. Elsevier.

Duranton, G. and Puga, D. (2004). Micro-Foundations of Urban Agglomeration Economies. In *Handbook of Regional and Urban Economics*, volume 4, pages 2063–2117. Elsevier.

Edlund, L. (2005). Sex and the City. *Scandinavian Journal of Economics*, 107(1):25–44.

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

Elvery, J. A. (2009). The Impact of Enterprise Zones on Resident Employment: An Evaluation of the Enterprise Zone Programs of California and Florida. *Economic Development Quarterly*, 23(1):44–59.

Fernández, R. and Rogerson, R. (2001). Sorting and Long-Run Inequality. *The Quarterly Journal of Economics*, 116(4):1305–1341.

Glaeser, E. L. and Gottlieb, J. D. (2008). The Economics of Place-Making Policies. *Brookings Papers on Economic Activity*, pages 155–239.

Glaeser, E. L. and Mare, D. C. (2001). Cities and Skills. *Journal of Labor Economics*, 19(2):316–342.

Goodman-Bacon, A. (2021). Difference-in-differences with Variation in Treatment Timing. *Journal of Econometrics*, 225(2):254–277.

Greenwood, J., Guner, N., Kocharkov, G., and Santos, C. (2014). Marry Your Like: Assortative Mating and Income Inequality. *American Economic Review*, 104(5):348–53.

Grossbard-Schectman, S. and Grossbard-Shechtman, S. (1993). *On the Economics of Marriage*. Routledge.

Ham, J. C., Swenson, C., İmrohoroğlu, A., and Song, H. (2011). Government Programs Can Improve Local Labor Markets: Evidence from State Enterprise Zones, Federal Empowerment Zones and Federal Enterprise Community. *Journal of Public Economics*, 95(7-8):779–797.

Hamilton, G. and Siow, A. (2007). Class, Gender and Marriage. *Review of Economic Dynamics*, 10(4):549–575.

Holmes, T. J. (1999). Localization of Industry and Vertical Disintegration. *Review of Economics and Statistics*, 81(2):314–325.

Hryshko, D., Juhn, C., and McCue, K. (2017). Trends in Earnings Inequality and Earnings Instability Among Us Couples: How Important Is Assortative Matching? *Labour Economics*, 48:168–182.

Imbruno, M. (2016). China and WTO Liberalization: Imports, Tariffs and Non-Tariff Barriers. *China Economic Review*, 38:222–237.

Lin, M.-J., Liu, J.-T., and Qian, N. (2014). More Missing Women, Fewer Dying Girls: The Impact of Sex-Selective Abortion on Sex At Birth and Relative Female Mortality in Taiwan. *Journal of the European Economic Association*, 12(4):899–926.

Liu, Q. and Ma, H. (2020). Trade Policy Uncertainty and Innovation: Firm Level Evidence from China's WTO Accession. *Journal of International Economics*, 127:103387.

Lu, Y., Wang, J., and Zhu, L. (2019). Place-based Policies, Creation, and Agglomeration Economies: Evidence from China's Economic Zone Program. *American Economic Journal: Economic Policy*, 11(3):325–60.

Lu, Y. and Yu, L. (2015). Trade Liberalization and Markup Dispersion: Evidence from China's WTO Accession. *American Economic Journal: Applied Economics*, 7(4):221–53.

Lynch, D. and Zax, J. S. (2011). Incidence and Substitution in Enterprise Zone Programs: The Case of Colorado. *Public Finance Review*, 39(2):226–255.

Ma, S., Xu, G., and Zhang, H. (2019). Input Tariff Liberalisation and Servitisation in Manufacturing: Evidence from China's WTO Accession. *The World Economy*, 42(12):3569–3587.

Marshall, A. (1890). *Principles of Economics*. Mcmillan, London.

Moretti, E. (2004). Workers' Education, Spillovers, and Productivity: Evidence from Plant-Level Production Functions. *American Economic Review*, 94(3):656–690.

Neumark, D. and Kolko, J. (2010). Do Enterprise Zones Create Jobs? Evidence from California's Enterprise Zone Program. *Journal of Urban Economics*, 68(1):1–19.

Neumark, D. and Simpson, H. (2015). Place-based Policies. In *Handbook of Regional and Urban Economics*, volume 5, pages 1197–1287. Elsevier.

Pencavel, J. (1998). Assortative Mating by Schooling and the Work Behavior of Wives and Husbands. *The American Economic Review*, 88(2):326–329.

Qian, N. (2008). Missing Women and the Price of Tea in China: The Effect of Sex-Specific Earnings on Sex Imbalance. *The Quarterly Journal of Economics*, 123(3):1251–1285.

Qian, Y. and Qian, Z. (2017). Assortative Mating by Education and Hukou in Shanghai. *Chinese Sociological Review*, 49(3):239–262.

Qin, B. and Zhang, Y. (2014). Note on Urbanization in China: Urban Definitions and Census Data. *China Economic Review*, 30:495–502.

Rosenthal, S. S. and Strange, W. C. (2004). Evidence on the Nature and Sources of Agglomeration Economies. In *Handbook of Regional and Urban Economics*, pages 2119–2177. Elsevier.

Sant'Anna, P. H. and Zhao, J. (2020). Doubly Robust Difference-in-differences Estimators. *Journal of Econometrics*, 219(1):101–122.

Shirk, S. L. et al. (1993). *The Political Logic of Economic Reform in China*, volume 24. Univ. of California Press.

Singh, S. and Samara, R. (1996). Early Marriage Among Women in Developing Countries. *International Family Planning Perspectives*, pages 148–175.

Wang, J. (2013). The Economic Impact of Special Economic Zones: Evidence from Chinese Municipalities. *Journal of Development Economics*, 101:133–147.

Xu, C. (2011). The Fundamental Institutions of China's Reforms and Development. *Journal of Economic Literature*, 49(4):1076–1151.

Yu, Y., Gong, L., and Yi, J. (2020). Globalization Raises Intergenerational Inequality Transmission in Chinese Villages. *Working Paper*.

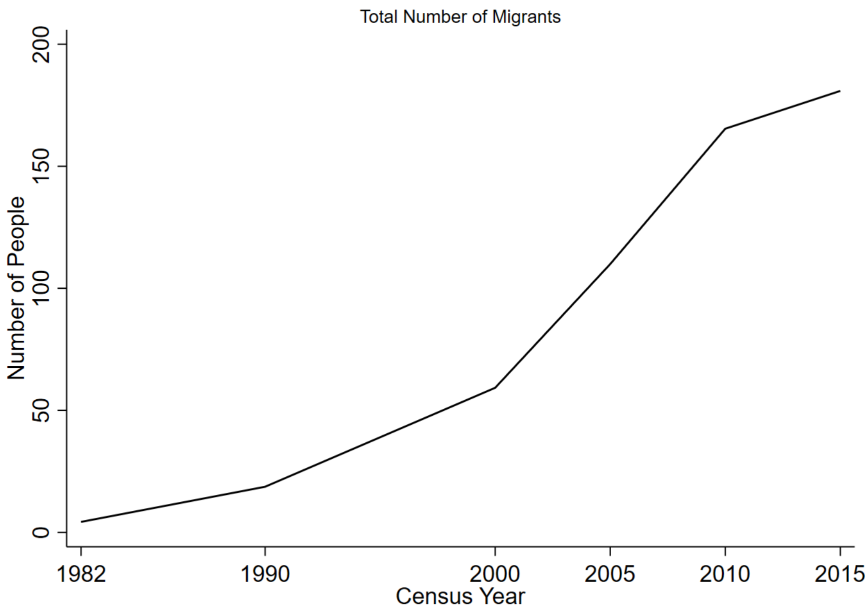


Figure 3.1: Total Number of Migrants

Notes: This figure shows the total number of migrants who are above 16 in each census year. Migrants are defined as inter-county movers whose residential address is different from his/her Hukou address. The numbers are counted in 1 million.

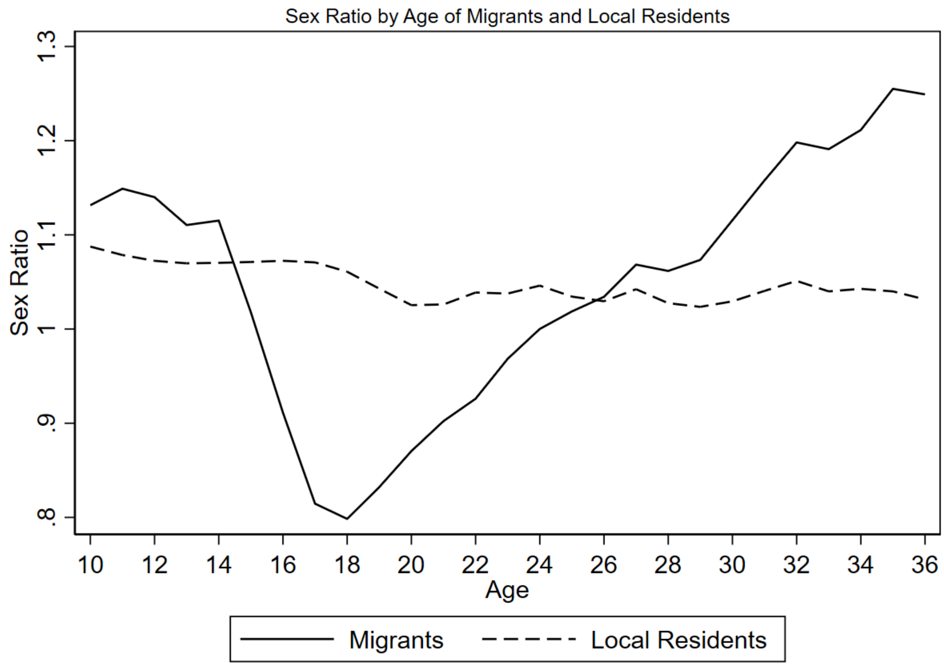


Figure 3.2: Sex Ratio by Age of Migrants and Local Residents

Notes: The sex ratio is defined as total number of males divided by number of females in the National Population Census of China.

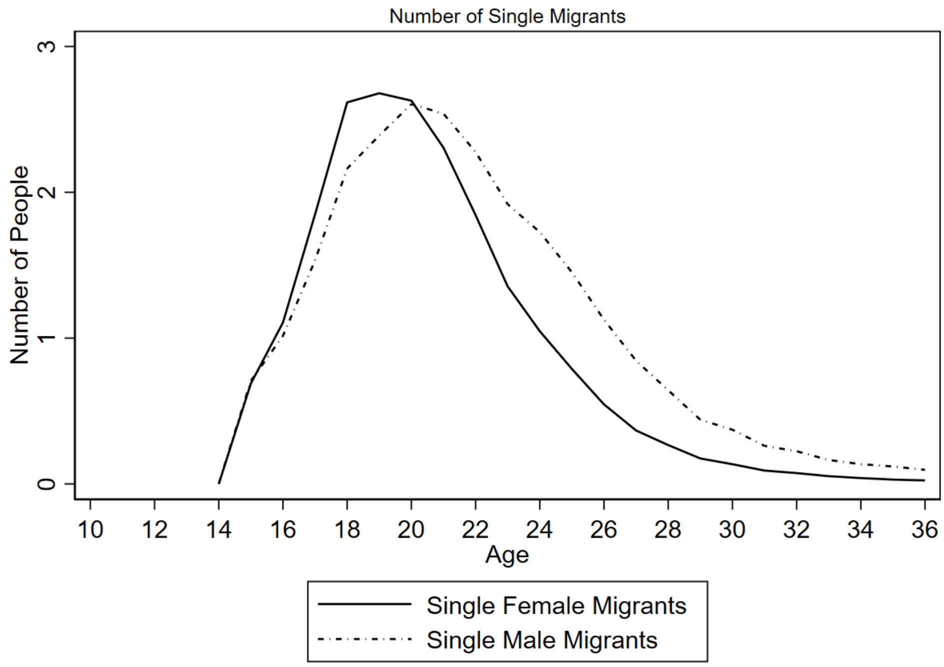


Figure 3.3: Number of Single Migrants

Notes: Number of migrants are counted in 10,000. Data are obtained by aggregating number of migrants across all six waves of the National Population Census of China.

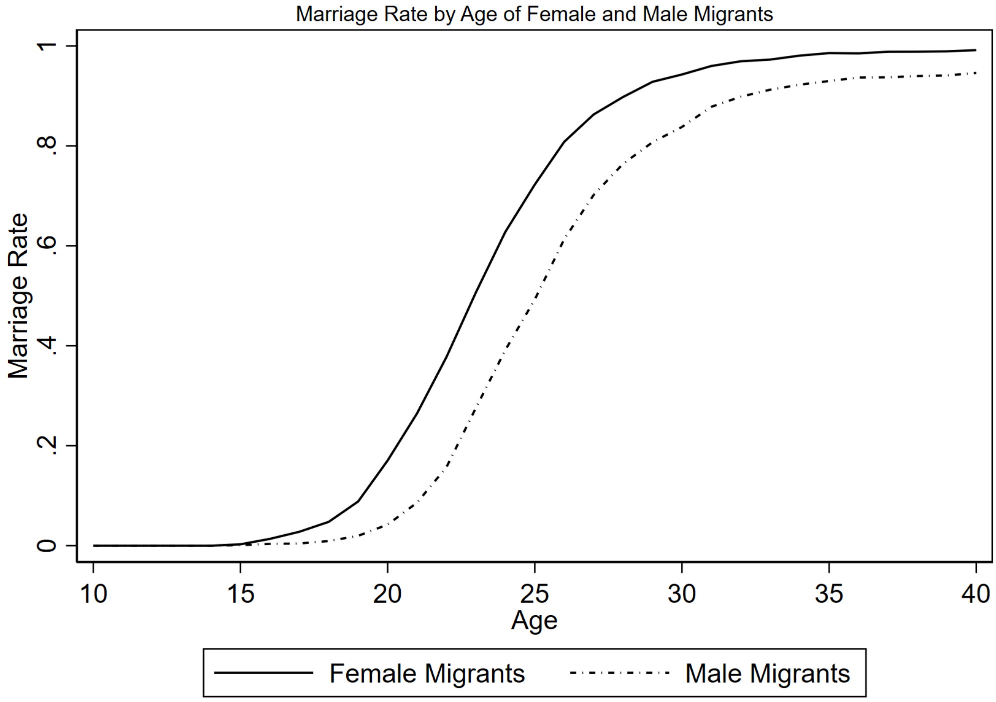


Figure 3.4: Marriage Rate by Age of Female and Male Migrants

Notes: Data are obtained by aggregating marital status of migrants across all six waves of the National Population Census of China.

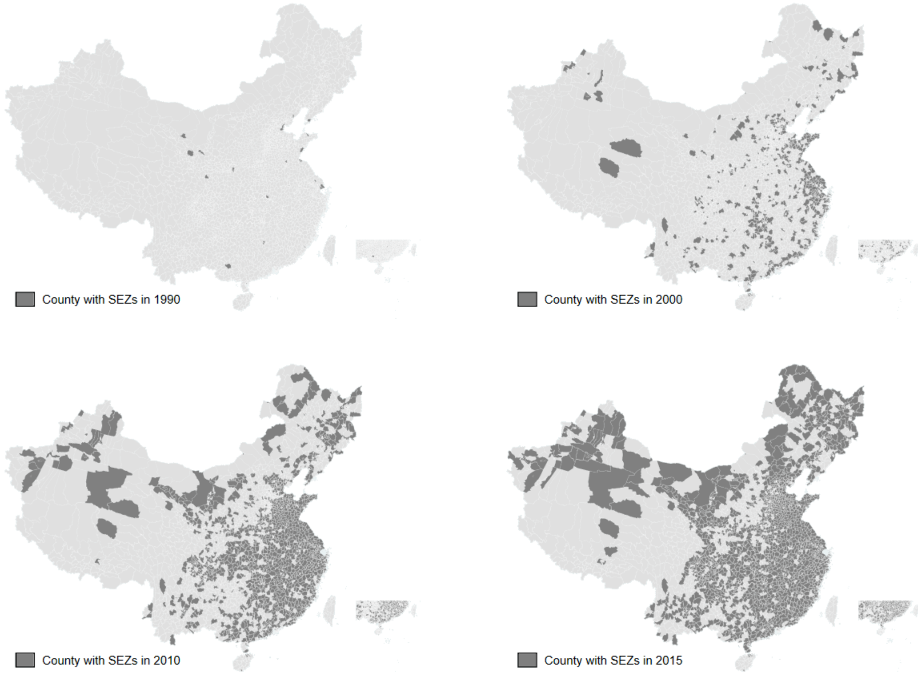


Figure 3.5: Establishment of SEZs

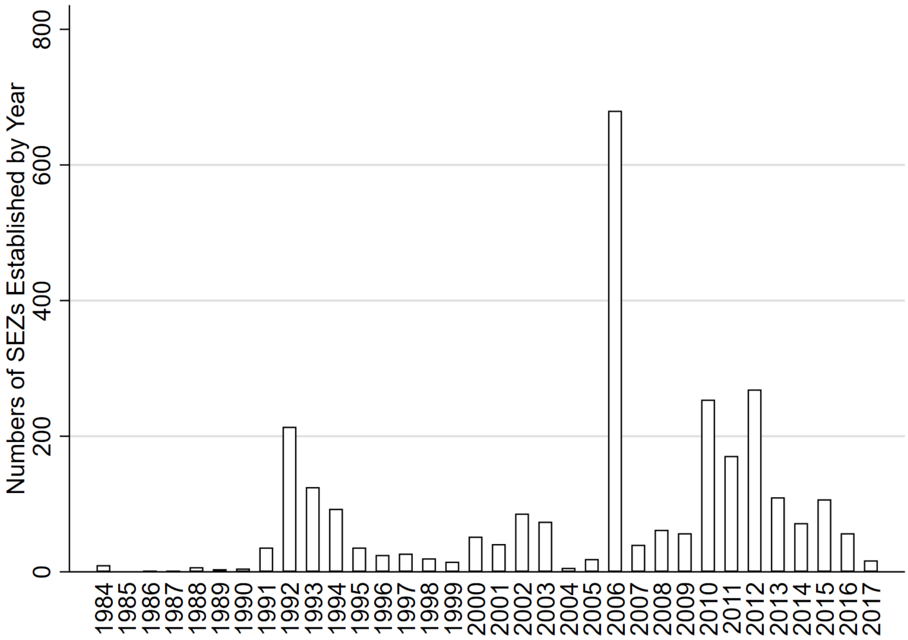


Figure 3.6: Numbers of SEZs established by Year

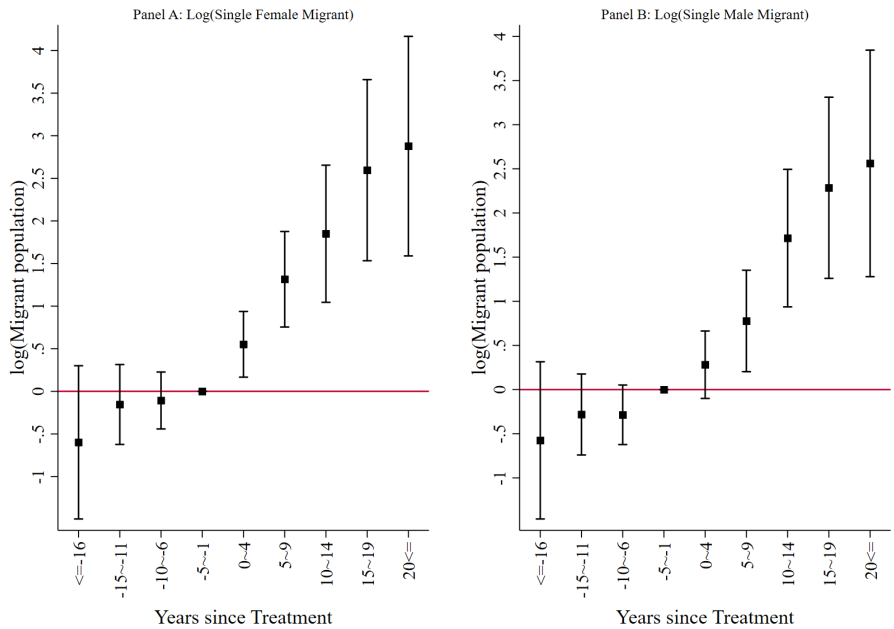


Figure 3.7: Time-Varying Effects of SEZ on Migrants -County Level (TWFE)

Notes: This figure reports the estimated effects of SEZ on innbound single female (Panel A) and male (Panel B) migrants who are aged between 16 and 25. The effects of years relative to SEZ establishment are plotted, along with bars representing the 95% confidence intervals.

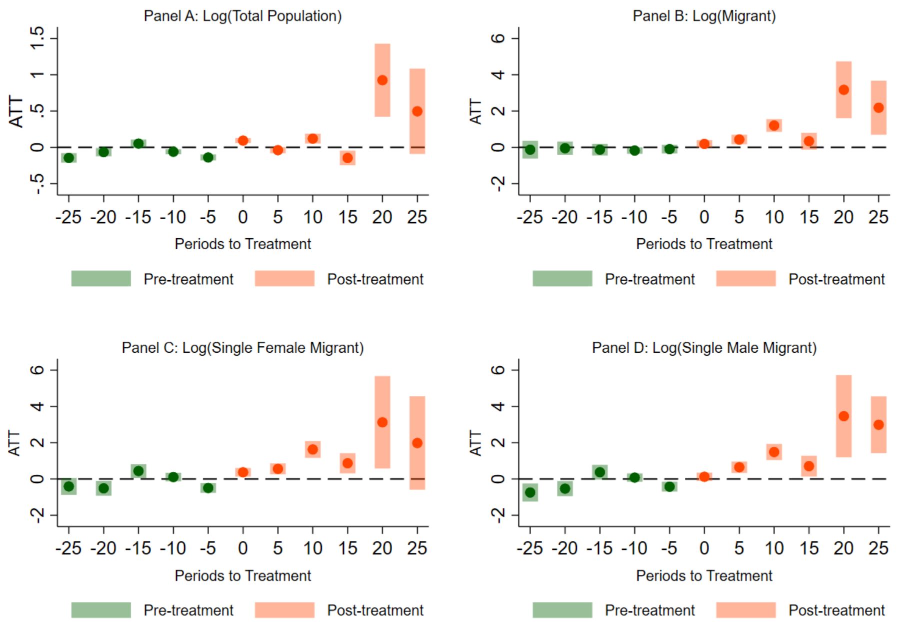


Figure 3.8: Time-Varying Effects of SEZ on Population - County Level Analysis (DiD)

Notes: The dynamic ATT's are plotted using the staggered DiD model, based on Callaway and Sant'Anna (2021) and Sant'Anna, Pedro, and Zhao (2020). The control group consists of the counties that were never-treated during our sample period. Panel A reports the effects on total population. Panel B, C and D report the effects on migrants aged between 16 and 25.

Table 3.1: Summary of Statistics – County Level

	Mean	SD	Min	Max	N
Treated Year (as of 2015)	2,003.65	7.63	1,984.00	2,015.00	1768
log(Treated Area) (as of 2015. Unit: ha)	956.53	1,533.37	5.30	31,337.84	1768
log(Population) in 1990	13.49	0.81	9.93	15.45	1338
log(Population) in 2000	13.56	0.66	10.71	15.34	1508
log(Population) in 2010	13.61	0.70	10.00	16.07	1702
log(Migrants) in 1990	8.14	2.37	0.00	14.00	1338
log(Migrants) in 2000	9.77	1.36	0.00	14.16	1508
log(Migrants) in 2010	10.94	1.62	0.00	15.38	1702
Sex ratio (Males to Females) in 1990	1.06	0.15	0.76	3.81	1338
Sex ratio (Males to Females) in 2000	1.05	0.05	0.81	1.36	1508
Sex ratio (Males to Females) in 2010	1.04	0.11	0.69	2.83	1702

Note: The sample includes all counties that ever had a SEZ established during our sample period.

Table 3.2: Average Effects of SEZ on Population and Inbound Migrants – County Level

	Total Population	All Migrants	Single Female Migrants	Single Male Migrants
Panel A	(1)	(2)	(3)	(4)
Treatment × After	0.031**	0.170*	0.399***	0.141
	(0.013)	(0.099)	(0.122)	(0.121)
Observations	9200	9200	9200	9200
Adjusted R^2	0.804	0.453	0.483	0.473
Mean of Dep. Variable	13.539	7.421	4.848	5.433
County FE	Yes	Yes	Yes	Yes
Province × Year FE	Yes	Yes	Yes	Yes
Panel B.1				
Treatment × After	0.036*	0.476***	0.686***	0.557***
	(0.019)	(0.106)	(0.126)	(0.127)
Method	DiD	DiD	DiD	DiD
Control Group	Not-yet	Not-yet	Not-yet	Not-yet
Panel B.2				
Treatment × After	0.033*	0.437***	0.669***	0.561***
	(0.020)	(0.105)	(0.125)	(0.127)
Method	DiD	DiD	DiD	DiD
Control Group	Never	Never	Never	Never

Note: Treatment status $\mathbb{1}$(SEZ) equals 1 starting from the year of SEZ establishment. Column (1) reports the effects of the SEZ establishment on total population. Column (2)–(4) reports the effects on inbound migrants who are aged between 16 and 25 (inclusive). Standard errors in parentheses are corrected for heteroskedasticity and clustered at the county level. Asterisks ***/**/* denote p<0.01, p<0.05, p<0.1.

Table 3.3: Average Effects of SEZ on Female Indicator – Individual Level

	Single Migrants	Single Migrants with Low Education	Single Migrants with High Education	Married Migrants
	(1)	(2)	(3)	(4)
Dependent Variable	$\mathbb{1}$(Female)	$\mathbb{1}$(Female)	$\mathbb{1}$(Female)	$\mathbb{1}$(Female)
$\mathbb{1}$(SEZ)	0.030**	0.031**	0.032	-0.011
	(0.013)	(0.014)	(0.022)	(0.012)
Observations	136311	96100	39910	48515
Adjusted R^2	0.061	0.080	0.077	0.071
Mean of dep. var	0.449	0.440	0.459	0.650
County FE	Yes	Yes	Yes	Yes
Province $\times$ Year FE	Yes	Yes	Yes	Yes
Age FE	Yes	Yes	Yes	Yes

Note: The dependent variable is $\mathbb{1}$(Female), which equals 1 if an individual is female and 0 otherwise. Individuals are located in counties that ever had SEZ during our sample period. Across all columns, we focus on individuals who are aged between 16 and 25 (inclusive) and are originally from rural Hukou. We use weights in the estimation to adjust for the different sampling sizes across census years. Standard errors in parentheses are corrected for heteroskedasticity and clustered at the county level. Asterisks ***/**/* denote p<0.01, p<0.05, p<0.1

Table 3.4: Average Effects of SEZ on Female Indicator with Ind and Occ FE – Individual Level

	Single Migrants	Single Migrants with Low Education	Single Migrants with High Education	Married Migrants
	(1)	(2)	(3)	(4)
Dependent Variable	𝟙(Female)	𝟙(Female)	𝟙(Female)	𝟙(Female)
𝟙(SEZ)	0.033***	0.033**	0.034	-0.013
	(0.012)	(0.014)	(0.022)	(0.010)
Observations	136310	96100	39909	48514
Adjusted R^2	0.100	0.136	0.104	0.265
Mean of dep. var	0.449	0.440	0.459	0.650
County FE	Yes	Yes	Yes	Yes
Province × Year FE	Yes	Yes	Yes	Yes
Age FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Occupation FE	Yes	Yes	Yes	Yes

Note: The dependent variable is 𝟙(Female), which equals 1 if an individual is female and 0 otherwise. Individuals are located in counties that ever had SEZ during our sample period. Across all columns, we focus on individuals who are aged between 16 and 25 (inclusive) and are originally from rural Hukou. We use weights in the estimation to adjust for the different sampling sizes across census years. Standard errors in parentheses are corrected for heteroskedasticity and clustered at the county level. Asterisks ***/**/* denote p<0.01, p<0.05, p<0.1

Table 3.5: Average Effects of the WTO Accession on Female Indicator – Individual Level

	Single Migrants	Single Migrants with Low Education	Single Migrants with High Education	Married Migrants
	(1)	(2)	(3)	(4)
Dependent Variable	$\mathbb{1}$(Female)	$\mathbb{1}$(Female)	$\mathbb{1}$(Female)	$\mathbb{1}$(Female)
WTOshock × Post	0.014**	0.025***	0.007	0.008
	(0.006)	(0.009)	(0.009)	(0.009)
Observations	90725	72819	17779	18439
Adjusted R^2	0.103	0.136	0.116	0.117
Mean of dep. var	0.468	0.494	0.427	0.586
County FE	Yes	Yes	Yes	Yes
Province × Year FE	Yes	Yes	Yes	Yes
Age FE	Yes	Yes	Yes	Yes

Note: The dependent variable is $\mathbb{1}$(Female), which equals 1 if an individual is female and 0 otherwise. Years 2000 and 2015 are used. Across all columns, we focus on individuals who are aged between 16 and 25 (inclusive) and are originally from rural Hukou. We use weights in the estimation to adjust for the different sampling sizes across census years. Standard errors in parentheses are corrected for heteroskedasticity and clustered at the county level. Asterisks ***/**/* denote p<0.01, p<0.05, p<0.1

Table 3.6: Average Effects of the WTO Accession on Female Indicator – with Ind and Occ FE

	Single Migrants	Single Migrants with Low Education	Single Migrants with High Education	Married Migrants
	(1)	(2)	(3)	(4)
Dependent Variable	$\mathbb{1}$(Female)	$\mathbb{1}$(Female)	$\mathbb{1}$(Female)	$\mathbb{1}$(Female)
WTOshock $\times$ Post	0.012**	0.024***	0.007	-0.001
	(0.006)	(0.009)	(0.009)	(0.009)
Observations	90725	72819	17779	18439
Adjusted R^2	0.138	0.182	0.142	0.280
Mean of dep. var	0.468	0.494	0.427	0.586
County FE	Yes	Yes	Yes	Yes
Province $\times$ Year FE	Yes	Yes	Yes	Yes
Age FE	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes
Occupation FE	Yes	Yes	Yes	Yes

Note: The dependent variable is $\mathbb{1}$(Female), which equals 1 if an individual is female and 0 otherwise. Years 2000 and 2015 are used. Across all columns, we focus on individuals who are aged between 16 and 25 (inclusive) and are originally from rural Hukou. We use weights in the estimation to adjust for the different sampling sizes across census years. Standard errors in parentheses are corrected for heteroskedasticity and clustered at the county level. Asterisks ***/**/* denote p<0.01, p<0.05, p<0.1

Table 3.7: Average Effects of SEZ on Marriage – Individual Level

	(1) $\mathbb{1}$(Married)	(2) $\mathbb{1}$(Married)	(3) $\mathbb{1}$(Married to Local)	(4) $\mathbb{1}$(Married to Local)
$\mathbb{1}$(SEZ) × $\mathbb{1}$(Female)	0.114***	0.133***	0.004**	0.003
	(0.005)	(0.003)	(0.002)	(0.002)
$\mathbb{1}$(SEZ)	-0.064***	-0.076***	-0.006	-0.005
	(0.009)	(0.006)	(0.007)	(0.007)
Observations	283977	283977	85466	85466
Adjusted R^2	0.080	0.470	0.276	0.277
Mean of dep. var	0.421	0.421	0.054	0.054
County FE	Yes	Yes	Yes	Yes
Province × Year FE	Yes	Yes	Yes	Yes
Age FE	No	Yes	No	Yes

Note: This table reports the effects of SEZ on marriage for rural-urban migrants. The dependent variable is $\mathbb{1}$(Married) for column (1) and (2), which equals 1 a migrant is married. The dependent variable is $\mathbb{1}$(Married to Local) for column (3) and (4), which equals 1 if a migrant is married to a local resident. Across all columns, we focus on migrants who are aged between 16 and 30 (inclusive) and are originally from rural Hukou. We use weights in the estimation to adjust for the different sampling sizes across census years. Standard errors in parentheses are corrected for heteroskedasticity and clustered at the county level. Asterisks ***/**/* denote p<0.01, p<0.05, p<0.1

Appendix

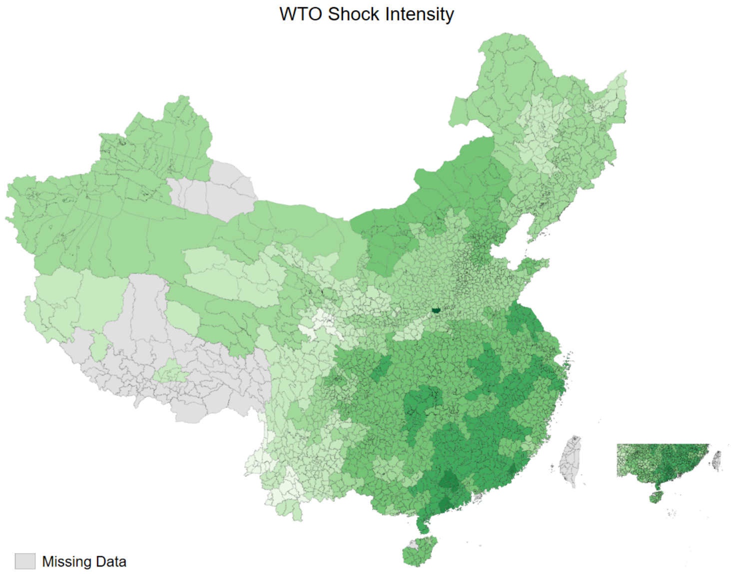


Figure A1: Exposure to China's Accession to the WTO

Notes: This figure plots the exposure to China's accession to the WTO, constructed using Equation (3.5). Darker colors represent stronger push force to migrant workers, which makes them have lager tendency to leave.

Figure A2: Origin Counties

Notes: This figure plots the origin counties used in Table 3.6.