Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

The Landscape and Dynamics of Private Supplementary Tutoring in Urban China

Jian Huang
Southwestern University of Finance and Economics
Zhaoyuan Chen
Southwestern University of Finance and Economics
Nike Zhao
Shanghai University of Finance and Economics
John Sessions
john.sessions@newcastle.ac.uk
Newcastle University

Article

Keywords: Public Education Spending, Educational inequality, Gini Coefficient

Posted Date: May 19th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9440328/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

The Landscape and Dynamics of Private Supplementary Tutoring in Urban China

Abstract: Using nationwide point-of-interest data on private supplementary tutoring (PST) institutions, we map the spatial landscape of PST and examines its dynamics across urban China from 2015 to 2018. Our results reveal that PST is strongly agglomerated in major cities and that high-density areas not only persist, but also intensify over time. At the school-district level, the distribution of PST is highly unequal, with Gini coefficients exceeding 0.8 in both years. We further show that the presence of PST institutions in adjacent districts substantially increases the likelihood of new establishments in previously uncovered areas, and that such spillover effects tend to be larger when cities experience rapid growth in the share of one-child families. Overall, the findings indicate a contagious pattern of PST expansion driven by status-oriented educational demand in competitive environments.

Key Words: Public Education Spending, Educational inequality, Gini Coefficient
JEL Classifications: J31, J32, J33, M12.

22ND April 2026

1. Introduction

Private supplementary tutoring (PST) - also termed 'shadow education' - has expanded rapidly worldwide [Bray (2017), Yamato and Zhang (2017), Byun et al. (2018), Jansen et al. (2023)]. PST provides tutoring closely aligned with the formal national curriculum and is widely recognized for enhancing students' academic performance and contributing to human capital accumulation [Dang and Rogers (2008), Kuan (2011), Byun and Park (2012)]. Indeed, the tendency of parents to evaluate their children's educational outcomes relative to others - known as 'status externalities' - can generate substantial demand for PST and intensify competition in the tutoring market [Anderson and Kohler (2013), Kim et al. (2024), Dai and Zhou (2025)]. Such status-driven competition often places heavy financial pressure on households, exacerbates educational inequalities, and can lead to inefficient over-investments in human capital [Kim and Lee (2010), Chung and Lee (2017), Oh et al. 2020), Kumar and Chodhury (2021)].

Despite its growing significance, comprehensive large-scale evidence on PST remains scarce. Operating largely outside of the formal education system, it remains loosely institutionalized and is often omitted from official statistics, leaving most existing studies reliant on household surveys or self-reported questionnaires. Consequently, the scale and dynamics of PST remain moot, and the influence of status-driven demand has yet to be substantiated with robust empirical evidence. We address this gap by mapping the PST landscape and examining its spatial dynamics in urban China.

China's formal education system is highly stratified and selective, motivating parents to evaluate their children's performance relative to peers [Zhang and Bray (2016), Lu et al.

(2022)]. Consequently, PST has become a common strategy for parents who fear their children may 'fall behind at the starting line' in educational competition [Li and Xie (2020), Chen et al. (2021), Chen et al. (2022)] . The cultural and demographic contexts of the country may further intensify educational competition and amplify status-driven demand for PST. Confucian traditions, which link academic achievement to social standing and emphasize parental responsibility for children's educational success, have been associated with heightened parental concern over academic performance [Chen et al. (2021)]. Demographic shifts may also play a role, as shrinking family sizes following the one-child policy have concentrated expectations and resources on each child's educational prospects [Zhang and Zhu (2008), Rosenzweig and Zhang (2009)].

Overall, China's educational and cultural structures are grounded in meritocratic norms that generate enduring status anxiety, driving parental investment in private tutoring. Against this backdrop, we investigate the distribution of PST in urban China and examines how status-driven demand contributes to its expansion. To achieve this goal, we use point-of-interest (POI) data on tutoring institutions for 2015 and 2018, a period characterized by rapid and largely unregulated expansion of the PST industry. Unlike studies relying on self-reported information, our approach provides supply-side evidence that captures the intensity of PST activity across districts, cities, and the nation. To our knowledge, this is the first study to conduct a comprehensive examination of tutoring institutions across multiple spatial scales. Our findings illuminate the scope and underlying drivers of China's shadow education and underscore its implications for educational equity. Our analysis may also offer insights into the broader 'education fever' and the proliferation of private tutoring in other East Asian societies shaped

by Confucian cultural traditions and highly competitive education systems [Kim and Lee (2010), Stevenson and Baker (1992), Marginson (2011), Kwok (2024)].

2. Results

We utilize POI data from 'Amap' -also known as '*Gaode Maps'* - China's leading digital mapping and navigation platform, to examine the spatial distribution of PST institutions in 2015 and 2018. Natural Language Processing (NLP) methods are applied to classify and verify information on individual institutions. Our dataset includes 240613 tutoring institutions in 2015, of which 116167 offered tutoring in academic subjects. By 2018, this number had increased to 481565, with 233089 institutions providing academic tutoring. We focus on academic tutoring, which is more directly associated with educational competition and thus relevant for understanding the role of status externalities in shaping family educational investment.

Figure 1 illustrates the distribution of PST institutions across 282 Chinese cities. We exclude non-urban prefectural units from our analysis and focus instead on prefecture-level cities and centrally administered municipalities. In these latter, the total number of PST institutions rose from 113472 in 2015 (98.13% of the national total) to 227363 in 2018 (97.82% of the national total). City-level counts ranged from 10 (Lincang) to 5134 (Beijing) in 2015 and from 34 (Lincang) to 7721 (Beijing) in 2018. These figures reveal large differences in local tutoring supply, as well as a rapid expansion of PST institutions across almost all cities between 2015 and 2018, with the national total doubling in just three years. This period marks the industrialization of PST with large tutoring companies scaling up their operations and many individual tutors entering the marketplace [Zhang and Bray (2018), Liu, (2018)].

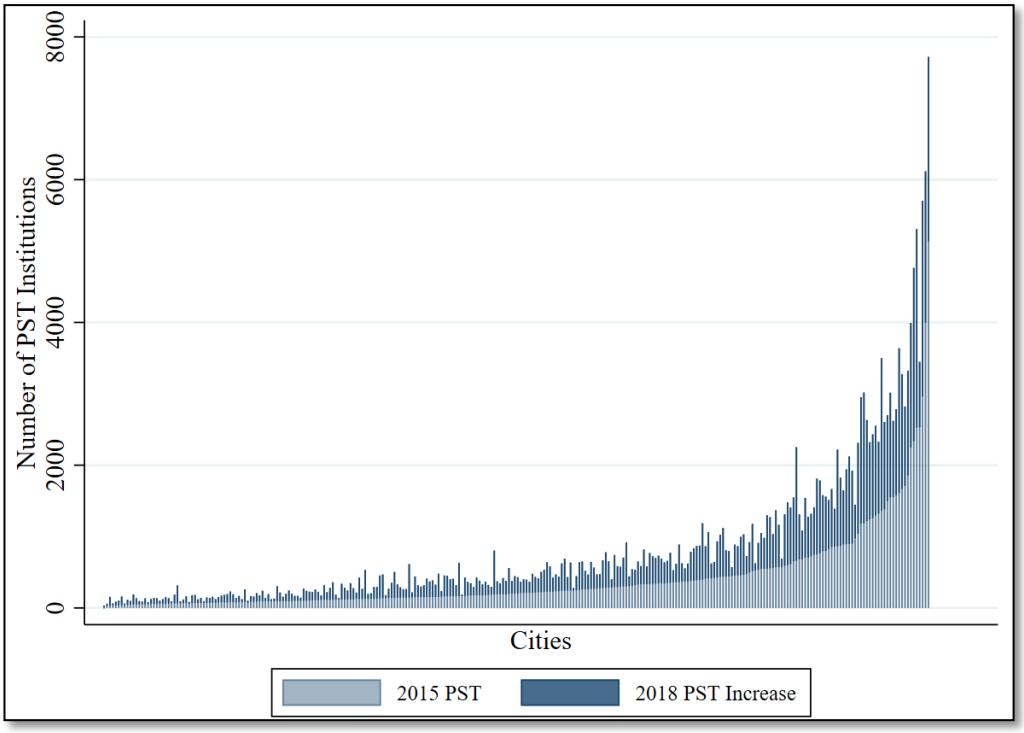


Figure 1: Number of PST Institutions by Cities, 2015 and 2018[1]

Macro-Spatial Analysis

We begin with a macro-spatial analysis using kernel density estimation to explore the landscape and dynamics of PST in urban China. Kernel density maps visualizing the spatial distributions of PST institutions in 2015 and 2018 are presented in Figure 2a and Figure 2b.

Figure 2a reveals that high-PST areas (intensity > 0.1) are concentrated in eastern and southeastern cities, while they are rarely present in western regions. In particular, there are strong geographic agglomerations in three clustering regions: (i) the Beijing–Tianjin (B–T) region; (ii) the Jiangsu–Zhejiang–Shanghai (J–Z–H) region; and (iii) the Guangzhou–Shenzhen (G–S) region. These are among the most economically developed urban regions in China, jointly contributing 30.96% of national GDP and comprising 15.61% of the total population in 2015.

[1] Cities in Figure 1 are arranged in ascending order by their number of PST institutions in 2015. Stacked bars display the 2015 counts (base) and the subsequent increases between 2015 and 2018, with the total for 2018 represented by the full bar height; $N = 282$.

Figure 2a also highlights cities with top-ranked high schools, which largely coincide with areas of high PST density – see Appendix A6. This spatial overlap reflects the close alignment between elite public schooling and private tutoring markets, as top-ranked high schools are predominantly public institutions. In China's education system, general upper-secondary education is widely regarded as preparatory for higher education, and admission to top-ranked high schools confers significant advantages in accessing leading universities. Consequently, local educational competition carries particularly high stakes in cities with top-ranked high schools, prompting greater engagement in PST. It is also important to note that in metropolitan areas, school quality is largely capitalized into property prices [Feng and Lu (2013), Chan et al. (2020)].This allows affluent urban families to invest in high-quality public education through housing while simultaneously investing in shadow education.

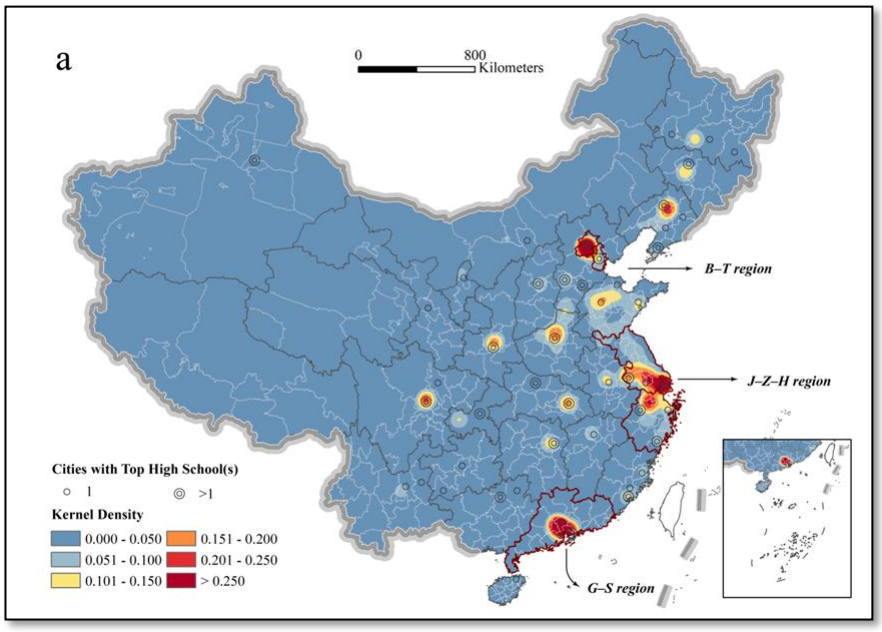


Figure 2a: Kernel Density Map of PST Institutions 2015

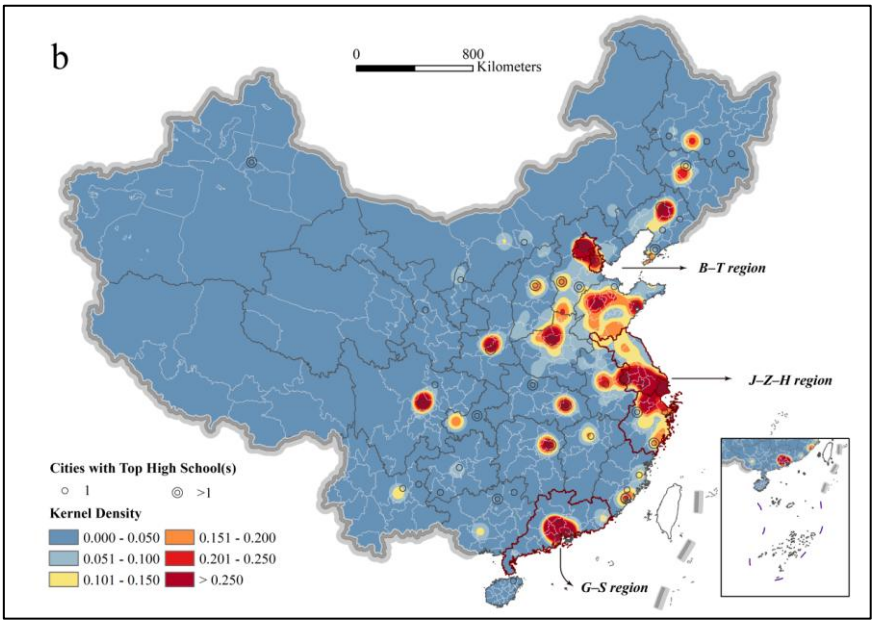


Figure 2b: Kernel Density Map of PST Institutions 2018

The 2018 density map set out in Figure 2b shows that the high PST areas identified in 2015 not only persist but also intensify over time. The three major clustering regions exhibit markedly stronger agglomeration, with their high-density zones becoming more pronounced and extensive. At the same time, cities hosting top-ranked high schools emerge as new hotspots, reflecting a surge in PST activity beyond the established clusters. These patterns suggest that educational competition in cities with elite schools has grown increasingly intense during this period.

A particularly pronounced expansion of PST is evident in Shandong Province, positioned between the B–T and J–Z–H clusters. As the birthplace of Confucius, Shandong has long upheld a strong cultural emphasis on scholarly achievement and educational advancement. Such an enduring academic tradition helps account for the sustained demand for private tutoring across the province. Anchored by the J–Z–H region and extending into Shandong, this part of eastern China now forms a contiguous belt of intensive PST activity, distinguishing itself as the region with the most intense educational competition in the country.

Figure 3 following provides a detailed illustration of the spatial diffusion of PST across cities, measured based on geographic adjacency. To construct a city-level measure of PST density, we calculated the mean of all raster cell values within each city's administrative boundary. This approach produces an indicator representing the kernel density of PST for each city.

Figure 3a presents a scatter plot of kernel density estimates for 282 cities. The horizontal axis represents the average kernel density of adjacent cities in 2015, while the vertical axis denotes the change in the city's own PST density from 2015 to 2018. Cities are categorized based on their initial PST levels in 2015, with those exhibiting a kernel density between 0 and 0.1 labeled as low-PST cities, and those exceeding 0.1 identified as intensive-PST cities. A strongly positive correlation is observed between the change in a city's PST density and the average PST density of neighboring cities ($r = 0.826$, $p < 0.001$).

Figure 3b shows the average increase in PST density among the three types of cities. Specifically, low-PST cities not adjacent to intensive-PST cities experienced an average increase of 0.023, while those adjacent to intensive-PST cities recorded a higher increase of 0.048. Cities with high PST density in 2015 saw the largest growth, with an average increase exceeding 0.173 over the 2015–2018 period.

Overall, Figure 3 reveals a clear pattern of spatial diffusion, as low-PST cities adjacent to high-PST areas experienced substantially greater increases in PST density than non-adjacent counterparts. Moreover, cities with high PST density in 2015 recorded the largest subsequent growth, consistent with the view that private tutoring can evolve into a self-reinforcing and potentially irreversible cycle in highly competitive environments [Bray (2017), Exley (2022)].

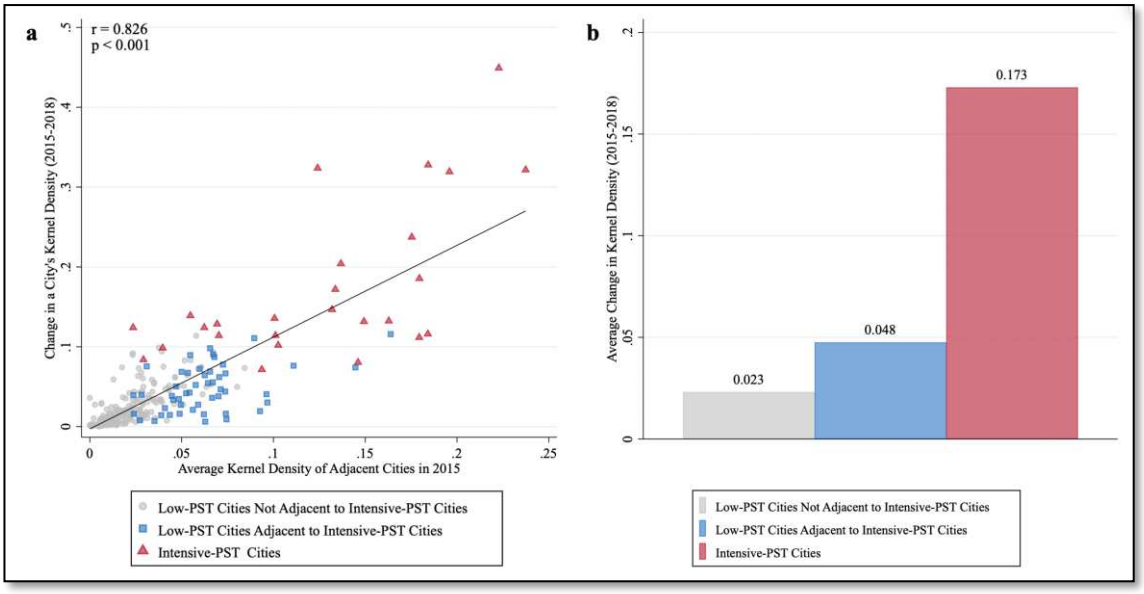


Figure 3: Spatial Diffusion of PST across Chinese Cities[2]

Local-Spatial Analysis

In this section we conduct a local-spatial analysis of PST distribution in 282 cities. Using 2015 POI data for primary and secondary schools, we define a school district as the area within a 500-meter circular buffer centered on each school. The 500-meter radius is based on the urban service standard for complete primary schools as specified in the *Code for Design of School Buildings* (GB 50099-2011). Detailed illustration of the spatial matching between schools and PST institutions is presented in Appendix A2.

Figure 4 displays the distribution of PST institutions across school districts. Figure 4**a** shows the share of districts by the number of PST institutions they host. In 2015, 75.88% of districts had no PST institution, while only 7.53% had more than 6. By 2018, the share of districts without any PST institution had declined to 69.07%, whereas the proportion with more than six increased to 12.63%.

[2] Figure 3a depicts the relationship between adjacent cities' PST density in 2015 and change in local PST density during 2015–2018. Figure 3b depicts the average increase in PST density by city type and proximity to intensive-PST cities. N =282.

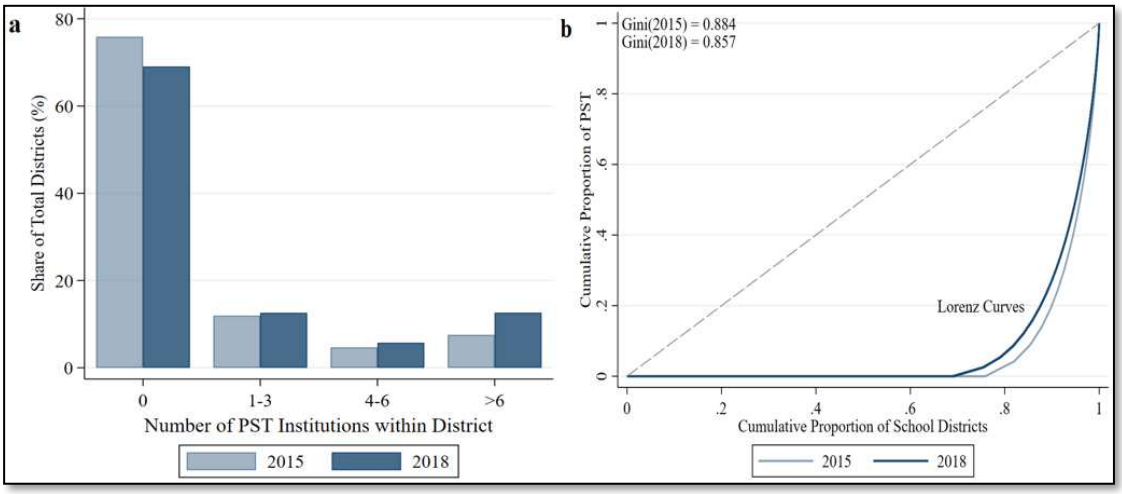


Figure 4: Distribution of PST across School Districts[3]

Figure 4b presents Lorenz curves and corresponding Gini coefficients that quantify the concentration of PST institutions across school districts. These measures are calculated using the total counts of PST institutions across districts rather than the number of unique institutions, as many are located within overlapping district boundaries. The Lorenz curves are sharply bowed away from the line of equality, indicating a highly uneven distribution. In 2015, the top 20%, 10%, and 1% of districts accounted for 97.1%, 79.8%, and 19.8% of all 347065mrecorded instances of PST institutions, respectively. A similarly high degree of concentration is observed in 2018, when the corresponding shares were 93.5%, 72.8%, and 16.8% of all 609056 recorded instances. The Gini coefficients derived from these frequency-based counts are 0.884 in 2015 and 0.857 in 2018. These results indicate that, despite the prevalence of PST nationwide, tutoring activities remain concentrated within a small fraction of school districts, reflecting enduring spatial inequalities in educational opportunities across urban China.

[3] Figure 4a depicts the share of school districts by the number of PST institutions hosted in 2015 and 2018. The horizontal axis groups school districts into four categories: 0, 1–3, 4–6, and more than 6 PST institutions. Figure 4b depicts Lorenz curves and corresponding Gini coefficients illustrating the concentration of PST institutions across school districts; the diagonal line represents the Line of Equality.

Figure 5 presents regression estimates quantifying the spillover effects of prior PST presence in adjacent districts - see also Appendix A4. The outcome variable, *Local PST Presence*, is a binary indicator equal to 1 if the district hosted at least one PST institution in 2018, and 0 otherwise. The key explanatory variable, *Adjacent PST Presence*, is defined similarly, taking the value 1 if any adjacent district contained PST institutions in 2015, and 0 otherwise. The coefficient on this latter variable captures the change in the likelihood of local PST presence associated with prior presence in adjacent districts.

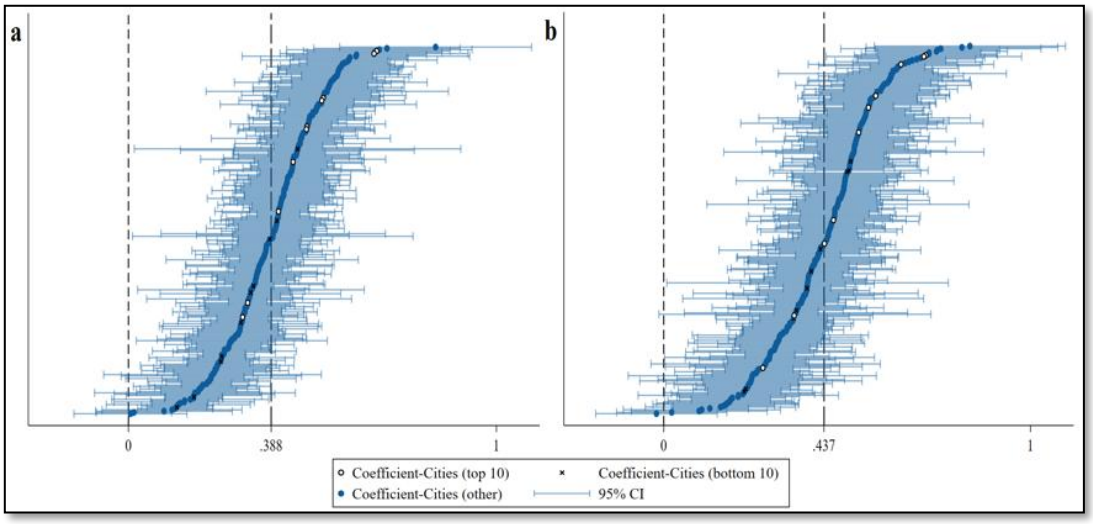


Figure 5: City-Specific Spillover Effects of Prior PST Presence in Adjacent School Districts[4]

Regressions are conducted separately for school districts within each city, so the resulting estimates reflect city-specific spillover effects. The first set of estimates, shown in Figure 5a, encompass all school districts, irrespective of their PST status in 2015. The second set, shown in Figure 5b, focus exclusively on districts without PST institutions in 2015. Estimated coefficients are displayed in ascending order for the 282 cities. In Figure 5a, they range from

[4] Figure 5a depicts City-specific estimates with 95% confidence intervals across all districts in each city. Figure5b depicts City-specific estimates with 95% confidence intervals for districts without local PST institution in 2015. In both panels, the 10 cities with the largest increases in the share of one-child families are highlighted with white circles, and the 10 cities with the largest declines are highlighted with black crosses.

0.005 to 0.835 with a mean of 0.388, and all city-level coefficients are positive. In Figure 5b, they range from –0.020 to 0.835 with a mean of 0.437, and only one estimate is negative. The estimation results from Figure 5b are particularly informative, as they reveal how PST institutions expanded into previously unserved areas. On average, districts without PST institutions in 2015 that were adjacent to those with such institutions were 43.7 percentage points more likely to host PST institutions in 2018 than districts without adjacent exposure.[5]

Figure 5 also incorporate demographic shifts in city-specific spillover effects. We highlight the 10 cities with the largest increases and the 10 cities with the largest declines in the share of one-child families between 2015 and 2018, defined as the proportion of households with exactly one child among all households with children aged 0–14. As shown in Figure 5a, estimates for the 10 cities with the largest increases in the share of one-child families are mainly concentrated in the upper half of the distribution, with 8 out of 10 in this range and 6 in the top quartile. By contrast, for the 10 cities with the largest declines, only two estimates fall in the upper half, and none are in the top quartile. A similar pattern is observed in Figure 5b. Among the 10 cities with the largest increases, seven estimates fall in the upper half of the distribution, with six in the top quartile. In contrast, for the 10 cities with the largest declines, only three estimates are in the upper half, and none reach the top quartile. Additional results in the Appendix A4 confirm that these effects tend to be more pronounced in cities with rapid growth

[5] We also present city-specific spillover estimates for non-academic PST in Figure A4 (Appendix). On average, these estimates are about 88 percent the size of the corresponding spillovers for academic PST, indicating that non-academic PST also generates substantial spillovers as parents seek to enhance their children's human capital beyond academic achievement. Nonetheless, these spillovers are less pronounced than those for academic PST, which is not unexpected given that parental concerns over "falling behind at the starting line" are more salient in academic competition and correspond to stronger status externalities.

in the share of one-child families.

Overall, Figure 5 offers evidence of status-driven PST competition at the district level. First, spillover effects are strongly positive in almost all Chinese cities - the presence of PST institutions in adjacent districts substantially increases the likelihood of their emergence in previously unserved area consistent with a status-driven spillover dynamic. Second, these effects tend to be more pronounced in cities experiencing rapid growth in the share of one-child families, where smaller family sizes and a stronger emphasis on child 'quality' amplify status externalities and intensify educational competition.

3. Discussion

Using POI data on PST institutions, we map the spatial landscape of shadow education and examines its dynamics across urban China. The macro-spatial analysis reveals strong agglomeration in major cities and clear evidence of spatial diffusion - cities adjacent to high-PST cities experienced significantly larger increases in tutoring intensity between 2015 and 2018. PST density is also markedly higher in cities hosting top-ranked high schools, and its growth over time is significantly greater in these cities. Our local-spatial analysis shows that PST is distributed unevenly, with 20% of school districts containing more than 90% of all PST institutions. The presence of PST institutions in adjacent districts substantially increases the likelihood of new establishments in previously uncovered areas, and such spillover effects tend to be more pronounced in cities experiencing rapid growth in the share of one-child families. Overall, our findings indicate that shadow education is both prevalent and socially contagious in urban China, driven by status-oriented educational demand in competitive urban

environments.

The tutoring boom driven by status-oriented demand has led to disproportionate educational investment, with far-reaching social and economic implications for the country. First, the expansion of PST exacerbates educational inequality. Tutoring institutions are concentrated in urban areas, particularly in cities with elite public schools. Affluent families in these metropolitan areas can reinforce their advantages through both stronger public education and greater spending on tutoring. This further entrenches socioeconomic status as a key determinant of educational opportunities [Zhang and Xie (2016), Liu and Bray (2017)]. Second, parental anxiety about children 'falling behind at the starting line' intensifies as tutoring becomes more widespread among peers, which leads to over-investment and inefficiencies in educational resource allocation in competitive environments [Chen et al. (2022), Zhang et al. (2025)]. Third, the intensification of PST competition exposes students to escalating academic pressure, resulting in heightened stress and anxiety among adolescents that has now emerged as a pressing social issue [Xu et al. (2020), Cao and Wang (2025)].

To address growing concerns over educational inequality and excessive academic competition, the Chinese government introduced stringent regulations in 2021 that banned for-profit tutoring in core curriculum subjects for grades 1–9. However, the ban has been highly controversial, as private tutoring inherently serves as a means of enhancing academic performance and accumulating human capital. The policy also generated substantial economic and employment consequences, affecting more than 200000 academic tutoring institutions nationwide.

Amid persistent demand for PST and a slowing economy, the regulations have recently

been relaxed. The policy debate now increasingly focuses on how to establish a more balanced and sustainable framework for regulating the private tutoring industry. Given that PST remains unevenly distributed across cities and districts and that its demand is largely status-driven, our findings suggest that mitigating China's tutoring fever requires coordinated efforts on both the supply and demand sides. On the supply side, policies should strengthen oversight of tutoring institutions and empower local governments to address the excessive concentration of PST institutions in highly competitive districts. On the demand side, broader reforms within the formal education system are needed to ease educational anxiety and promote more equitable access to learning opportunities.

References

Anderson, T. and Kohler, H. P. (2013). Education Fever and the East Asian Fertility Puzzle: A Case Study of Low Fertility in South Korea.' *Asian Population Studies*, 9, pp. 196-215.

Bray, M. (2017). 'Schooling and Its Supplements: Changing Global Patterns and Implications for Comparative Education.' *Comparative Education Review*, 61, pp. 469-491.

Byun, S.-Y. and Park, H. (20012). 'The Academic Success of East Asian American Youth: The Role of Shadow Education.' *Sociology of Education*, 85, pp. 40-60.

Byun, S.-Y., Chung, H. J. and Baker, D. P. (2018). 'Global Patterns of the Use of Shadow Education: Student, Family, and National Influences.' *Research in the Sociology of Education*, 20, pp. 71-105.

Cao, Y. and Wang, W. (2025). 'The Mental Health Costs of Shadow Education: The Duration of Shadow Education and its Depressive Effects on Chinese Adolescents.' *Behavioural Sciences*, 15.

Chan, J., Fang, X., Wang, Z., Zai, X. and Zhang, Q. (2020). 'Valuing Primary Schools in Urban China.' *Journal of Urban Economics*, 115, pp. 103-183.

Chen, G., Oubibi, M., Liang, A. and Zhou, Y. (2022). 'Parents' Educational Anxiety Under the "Double Reduction" Policy Based on the Family and Students' Personal Factors.' *Psychology Research and Behavior Management*, 15, pp. 2067-2082.

Chen, Y., Huang, R., Lu, Y. and Zhang, K. (2021). 'Education Fever in China: Children's Academic Performance and Parents' Life Satisfaction.' *Journal of Happiness Studies*, 22, pp. 927-954

Chung, K. and Lee, D. (2017). 'Inefficient Competition in Shadow-Education Investment.' *Journal of Economic Behavior and Organization*, 139, pp. 152-165.

Dai, C. and Zhou, Y. Rank. (2025). 'Peer Achievement, and Shadow Education: Evidence from Secondary School Students in China.' Economics of Education Review, 108, pp. 102-68.

Dang, H.-A. and Rogers, F. H. (2008). 'The Growing Phenomenon of Private Tutoring: Does It Deepen Human Capital, Widen Inequalities, or Waste Resources?' *The World Bank Research Observer*, 23, pp. 161-200.

Exley, S. (2022). 'Locked In: Understanding the 'Irreversibility' of Powerful Private Supplementary Tutoring Markets.' *Oxford Review of Education*, 48, pp. 78–94.

Feng, H. and Lu, M. (2013). 'School Quality and Housing Prices: Empirical Evidence from a Natural Experiment in Shanghai, China.' *Journal of Housing Economics*, 22(4), pp. 291-307.

Jansen, D., Elffers, L. and Jak, S. (2023). 'A Cross-National Exploration of Shadow Education Use by High- and Low-SES Families.' *International Studies in Sociology of Education*, 32, pp. 653-674.

Kim, S. and Lee, J.-H. (2010). 'Private Tutoring and Demand for Education in South Korea.' *Economic Development and Cultural Change*, 58, pp. 259-296

Kim, S., Tertilt, M. and Yum, M. (2024). 'Status Externalities in Education and Low Birth Rates in Korea.' *American Economic Review*, 114, pp. 1576-1611.

Kuan, P.-Y. (2011). 'Effects of Cram Schooling on Mathematics Performance: Evidence from Junior High Students in Taiwan.' *Comparative Education Review*, 55, pp. 342-368.

Kumar, I. and Chowdhury, I. R. (2021). 'Shadow Education in India: Participation and Socioeconomic Determinants.' *Journal of South Asian Development*, 16, pp. 244-272.

Kwok, P. L. Y. (2024). 'Theorizing Shadow Education and Academic Success in East Asia: Understanding the Meaning, Value and Use of Shadow Education by East Asian Students.' *Journal of Intercultural Studies*, 45, pp. 375-377

Li, H., Zhang, J. and Zhu, Y. (2008). 'The Quantity-Quality Trade-Off of Children in a Developing Country: Identification Using Chinese Twins.' *Demography*, 45, pp. 223-243

Li, W. and Xie, Y. (2020). 'The Influence of Family Background on Educational Expectations: A Comparative Study.' *Chinese Sociological Review*, 52, pp. 269-294

Liu, J. and Bray, M. (2017). 'Determinants of Demand for Private Supplementary Tutoring in China: Findings from a National Survey.' *Education Economics*, 25, pp. 205-218

Liu, J. (2018). 'Review of Regulatory Policies on Private Supplementary Tutoring in China.' *ECNU Review of Education*, 1, pp. 143-15.

Lu, W., Zhou, S. and Wei, Y. (2022). 'Government Policies and Unintended Consequences: Rising Demand for Private Supplementary Tutoring In China. International Journal of Educational Development 94, 102653

Marginson, S. (2011). 'Higher Education in East Asia and Singapore: Rise of the Confucian Model.' *Higher Education*, 61, pp. 587-611

Oh, B. C., Yeon, J. Y., Lee, H. S., Lee, D. W. and Park, E -C. (2020). 'Correlation Between Private Education Costs and Parental Depression in South Korea.' *BMC Public Health*, 20, p. 1-10.

Rosenzweig, M. R. and Zhang, J. (2009). 'Do Population Control Policies Induce mMre Human Capital Investment? Twins, Birth Weight and China's "One-Child" Policy.' *The Review of Economic Studies*, 76, pp. 1149-1174.

Stevenson, D. L. and Baker, D. P. Shadow Education and Allocation In Formal Schooling: Transition To University In Japan. American Journal Of Sociology 97, 1639-1657 (1992).

Xu, D. D., Cao, W., Wen, S., An, F., Che, W.,Bressington, D. T., Cheung, T., Ungvari G. S., and Xiang, Y. (2020). 'Prevalence of Depressive Symptoms in Primary School Students in China: A Systematic Review and Meta-Analysis.' *Journal of Affective Disorders*, 268, pp. 20-27

Yamato, Y. and Zhang, W. (2017). 'Changing Schooling, Changing Shadow: Shapes and Functions of Juku in Japan.' *Asia Pacific Journal of Education*, 37, pp. 329-343.

Zhang, L., Zhang, Y., Chen, Y. and Sun, J. (2000). 'Factors Influencing Parents' Educational Anxiety of Primary and Secondary School Students: Evidence from Parents in China.' BMC Public Health, 25, pp. 2-10.

Zhang, W. and Bray, M. A. (2020). 'Changing Environment of Urban Education: Historical and Spatial Analysis of Private Supplementary Tutoring in China.' *Environment and Urbanization*, 3, pp. 43-62.

Zhang, W. and Bray, M. (2016). 'Micro-Neoliberalism in China: Public-Private Interactions at the Confluence of Mainstream and Shadow Education.' *Journal of Education Policy*, 32, pp. 63-81

Zhang, Y. and Xie, Y. (2016). 'Family Background, Private Tutoring, and Children's Educational Performance in Contemporary China.' *Chinese Sociological Review*, 48, 64-82.

Appendix

A1. Data Sources

POI data points were retrieved from the Amap via its Application Programming Interface (API) for ad vanced place search (https://lbs.amap.com/api/webservice/guide/api-advanced/search).

Amap is one of the largest web mapping, navigation, and location-based service (LBS) platforms in China, providing information such as POI names, geographic coordinates, and category classifications. The POI categorization in Amap follows a hierarchical tree structure, consisting of three levels: top type, middle type, and bottom type. The corresponding classification codes are publicly available at https://lbs.amap.com/api/webservice/download. This study uses POI data of both PST institutions and schools, which fall under the top-level category "Science, Education, and Cultural Services".

The share of one-child families in 2015 and 2018 was calculated for each city using the 2015 National 1% Population Sample Survey and the 2020 National Population Census of China. The 2015 figures were derived from the sample survey, based on households with exactly one child among all households with children aged 0–14. The 2018 figures were obtained from the 2020 census, accessed via the National Bureau of Statistics–Peking University Research Data Center, and reflect households with exactly one child among all households with children aged 2–16, corresponding to the 0–14 age range in 2018.

A2. Identification of PST Institutions and Formal Schools

The Amap POI classification system lacks an exclusive subcategory for PST institutions and cannot effectively differentiate between secondary schools and vocational institutions. To address the problem, natural language processing (NLP) is applied to identify PST institutions as well as primary and secondary schools. Specifically, we trained a classification model using Bidirectional Encoder Representations from Transformers (BERT). BERT is a language representation model introduced by Google in 2018. Our BERT-based classification framework consists of pre-training procedure and fine-tuning procedure for PST classification.

During the pre-training stage, we used the bert-base-chinese model (https://huggingface.co/google-bert/bert-base-chinese), a Chinese-language pre-trained model built on the BERT architecture. It was trained primarily on Chinese Wikipedia to develop general language representations.

During the fine-tuning stage, BERT is initialized with pre-trained parameters, and all parameters are further adjusted using a manually labeled POI dataset to classify locations into specific types. The categories are:

a. Primary and General Secondary Schools – Public or private schools offering primary, lower-secondary, or general upper-secondary education (e.g., *No. 1 Middle School*). Vocational high schools, kindergartens, and universities are excluded;
b. Academic-oriented PST institutions – Private entities providing supplementary instruction in core subjects such as Chinese, mathematics, English, physics, chemistry, biology, geography, and history;
c. Non-academic-oriented PST institutions – Private entities offering training in arts and culture, including fine

arts, music, dance, calligraphy, and classical studies;

d. Adult vocational training institutions – Facilities delivering vocational training in areas such as auto repair, cosmetology, and hairdressing;

e. Others – that do not fall into the above categories or have unclear classification.

The fine-tuning procedure consisted of four steps. In Step 1, we extracted a 1% sample from the original 2015 POI dataset (1,486,632 entries) within the top category 'Science, Education, and Cultural Services' by selecting entries whose POI IDs ended with '00', yielding 14866 samples. Of these, 80% were randomly assigned to the training set and 20% to the test set.

In Step 2, the samples were annotated into five categories through a combination of GPT-assisted labeling and manual verification: primary and general secondary schools, academic-oriented PST, non-academic-oriented PST, adult tutoring institutions, and others.

In Step 3, we fine-tuned the bert-base-chinese model on the Google Colab platform. Classification outputs were generated by applying a softmax function to BERT embeddings, producing the most probable class and its associated confidence score. The BERT model was fine-tuned with a maximum token length of 256 and a batch size of 8 for 5 epochs. The learning rate was set to 2×10^{-5}, and the numerical stability parameter was 1×10^{-8}.

In Step 4, the trained model was used to predict categories for all POI data points for 2015 and 2018. To ensure prediction reliability, POIs with classification probabilities below 0.9 (71921 entities, 2.326% of the sample) were excluded from further analysis. The model achieved an accuracy between 90% and 95% on the test set, indicating the high reliability of BERT in POI classification.

We identified 240613 POI institutions in 2015, including 116167 offering tutoring in academic subjects. In 2018, the number of POI institutions increased to 481565, with 233089 offering tutoring in academic subjects. Additionally, we obtained POI data on 259020 primary and secondary schools in 2015. For comparison, the National Statistical Bulletin on Educational Development reported 256100 schools in 2015. The close alignment between our POI-based school counts and official statistics supports the reliability of our classification. The official bulletin is available at:

http://www.moe.gov.cn/srcsite/A03/s180/moe_633/201607/t20160706_270976.html.

We define a school district as the 500-meter area surrounding each school offering primary and/or secondary education, with every school serving as the district's focal point. PST institutions located within these buffer zones are assigned to their corresponding districts. Two districts are considered adjacent if the distance between their focal schools is less than 1000 meters; a district is classified as having no adjacent counterpart when all other schools are located more than 1000 meters away. Based on these definitions, we derive two district-level indicators that reflect the local and neighboring presence of PST institutions, as detailed below:

*Local Presence of PST Institution***.** A binary indicator equal to 1 if at least one PST institution is located within the school district, defined as being within a 500-meter radius of the district's focal school; and 0 otherwise.

*Adjacent Presence of PST Institution***.** A binary indicator equal to 1 if at least one PST institution is located in any school district adjacent to the local district. It equals 0 if all adjacent districts lack PST institutions, or if the local district has no adjacent districts (i.e. when all other schools are located more than 1000 meters from its focal school).

Figure A1 provides a specific example of these definitions. Red dots (*a, b, c*) denote schools and hollow dots (*x,*

y, z) represent PST institutions. Each blue circle indicates the 500-meter buffer around a school, defining the boundary of its corresponding school district. In this example, the district centered on School *a* contains PST institution *x*, and the district centered on School *c* contains PST institution *y* and *z*. The districts centered on Schools *b* and *c* are adjacent.

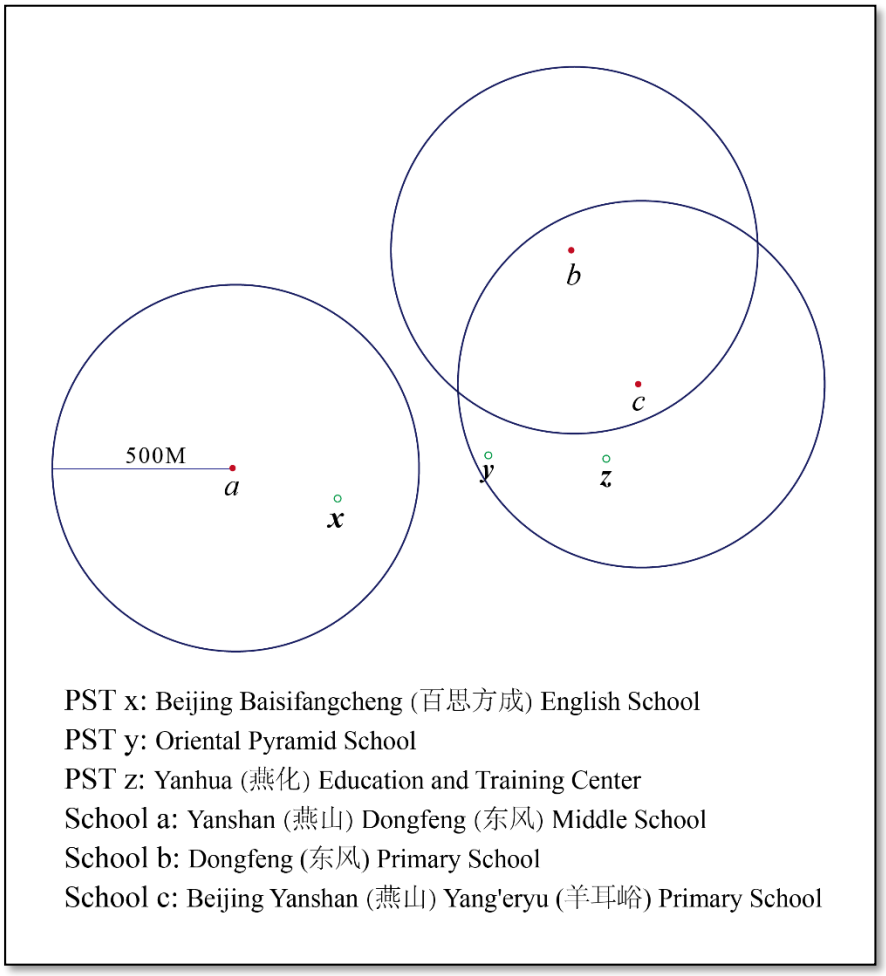


Figure A1: Visualizing School Districts, PST Presence, and Adjacency[6]

A3. Kernel Density Estimation

In the macro-spatial analysis, we applied a kernel density estimation (KDE) to examine the spatial distribution of PST institutions across urban China. KDE produces a continuous density surface of PST locations, which is partitioned into a grid of $1km \times 1km$ cells. The kernel density value for each cell aggregates the contributions of all PST locations within the specified bandwidth of the kernel function. We estimated the density at any location x, denoted by $f(x)$, by using the following formula:

$$f(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) \tag{1}$$

where n is the total number of PST institutions, x_i represents the location of a tutoring institution, h is the

[6] The district centered on School *a* contains PST *x*, and the district centered on School *c* contains PST *y* and *z*. The districts centered on Schools *b* and *c* are adjacent.

bandwidth that controls the smoothness of the estimate, and $K(\cdot)$ denotes a kernel function. For this estimation, we used the commonly applied quartic (biweight) kernel. The resulting kernel surface was converted into a raster grid of density values. Using this grid, city-level PST density was derived by extracting the kernel density values within each city's administrative boundary and computing their spatial mean.

A4. Regression Estimation

In the local-spatial analysis, we employ a linear probability regression model to quantify the spillovers of PST institutions. The dependent variable, $LocalPST_{ij}^{2018}$, indicates whether PST institutions are present in school district *j* within city *i* in 2018. The explanatory variables include the local presence of PST institutions in 2015, $LocalPST_{ij}^{2015}$, and the presence of PST institutions in adjacent districts in 2015, $NeighbourPST_{ij}^{2015}$. The model is specified as:

$$LocalPST_{ij}^{2018} = \beta_0 + \beta_1 LocalPST_{ij}^{2015} + \beta_2 NeighbourPST_{ij}^{2015} + \mu_i + \varepsilon_{ij} \quad (2)$$

where μ_i captures city fixed effects and ε_{ij} is the error term. In the city-specific regressions, coefficients β_1 and β_2 are allowed to vary across cities.

A5. Descriptive Statistics of PST Institutions

Our analysis covers 282 cities and 240254 school districts. Table A1 reports descriptive statistics on the distribution of PST institutions at both the city and district levels between 2015 and 2018. At the city level, the number of PST institutions ranged from 10 to 5134 in 2015 and from 34 to 7706 in 2018. The mean increased from 402.38 (*SD* = 580.68) in 2015 to 806.25 (*SD* = 1031.65) in 2018.

At the district level, the number of PST institutions ranged from 0 to 85 in 2015 and from 0 to 105 in 2018. The district-level average rose from 1.45 institutions (SD = 4.19) in 2015 to 2.54 institutions (SD = 6.47) in 2018. Local PST presence, measured as a binary indicator of whether a district contained at least one institution, also displayed spatial clustering. The share of districts with PST presence increased from 24.1% in 2015 to 30.9% in 2018.

Table A1. Descriptive Statistics

City-Level	*Observations.*	*Mean*	*S. Dev*	*Min*	*Max*
PST institution – 2015	282	402.383	580.679	10	5134
PST institution – 2018	282	806.252	1031.653	34	7706
District-Level	*Observations*	*Mean*	*Std*	*Min*	*Max*
PST institution – 2015	240254	1.445	4.186	0	85
PST institution – 2018	240254	2.535	6.473	0	105
Local PST presence 2015	240254	0.241	0.428	0	1
Local PST presence 2018	240254	0.309	0.462	0	1

A6. Top-Ranked High Schools in China

The list of top-ranked high schools in China in 2015 is provided by the 2015 Ranking of Chinese Hi gh Schools, compiled by the Yuantiku (猿题库) Evaluation and Research Center. The list includes 100 schools and is publicly available at: https://gaokao.eol.cn/news/201509/t20150908_1312976.shtml.

Figure A2 presents the distribution of cities hosting top-ranked high schools. These high schools are located in 49 cities, most of which are provincial capitals, sub-provincial cities, or centrally administered municipalities.



Figure A2: Distribution of Cities with Top-Ranked School(s)

A7. City-Specific Estimates of PST Spillover Effects

Demographic Shifts and City-Specific Spillover Effects

Supplementary Figure A3 shows average spillover effects of prior PST presence in adjacent school districts, grouped by changes in the share of one-child families. The magnitude of spillovers increases progressively from the large-decline category (< –0.1) to the stable (–0.1 to 0.1) and large-increase (> 0.1) categories.

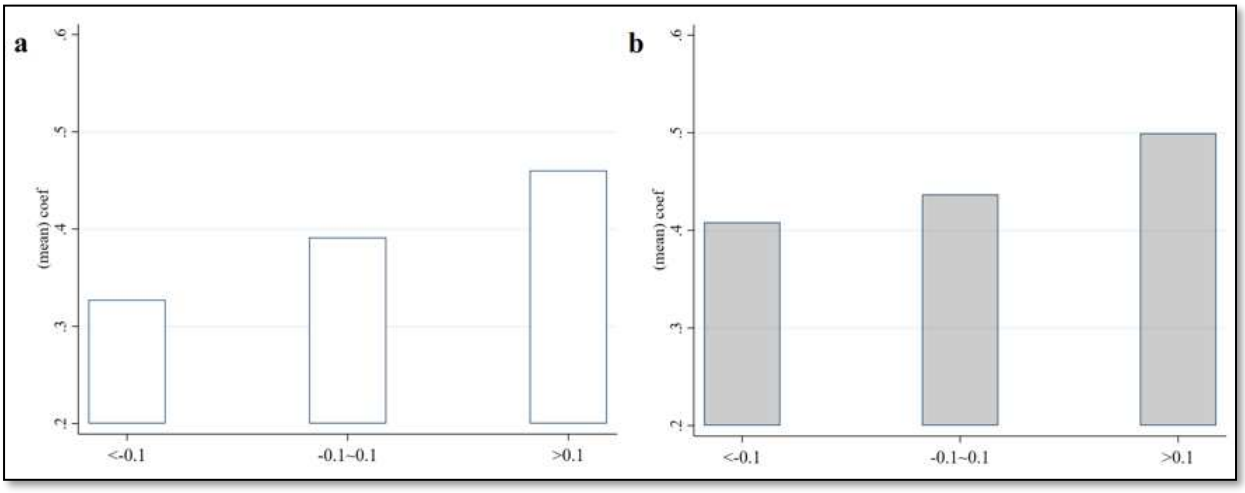


Figure A3: Spillover Effects of Prior PST Presence in Adjacent School Districts by Change in One-Child Family Share[7]

[7] Figure A3a presents city-specific mean spillover estimates across all districts within each city. Figure A3b presents city-

City-Specific Spillover Effects for Non-Academic PST

Figure A4 presents city-specific spillover estimates for non-academic PST. Figure A4a displays estimates for all school districts regardless of their 2015 PST status. Figure A4b focuses exclusively on districts without local PST institutions in 2015. The coefficients are arranged in ascending order, with confidence intervals shown for each city-specific estimate.

The coefficients in Figure A4a range from –0.011 to 0.684, with a mean of 0.34. A similar pattern appears in Figure A4b, where the coefficients range from –0.014 to 0.791 and the mean is 0.39. On average, these estimates are about 88% the magnitude of the corresponding spillovers for academic PST. Unlike the patterns observed for academic PST in Figure 5 of the main text, the spillover effects for non-academic PST do not exhibit a clear relationship with demographic shifts. These results suggest that parental concerns about "falling behind at the starting line" are less salient in non-academic competition.

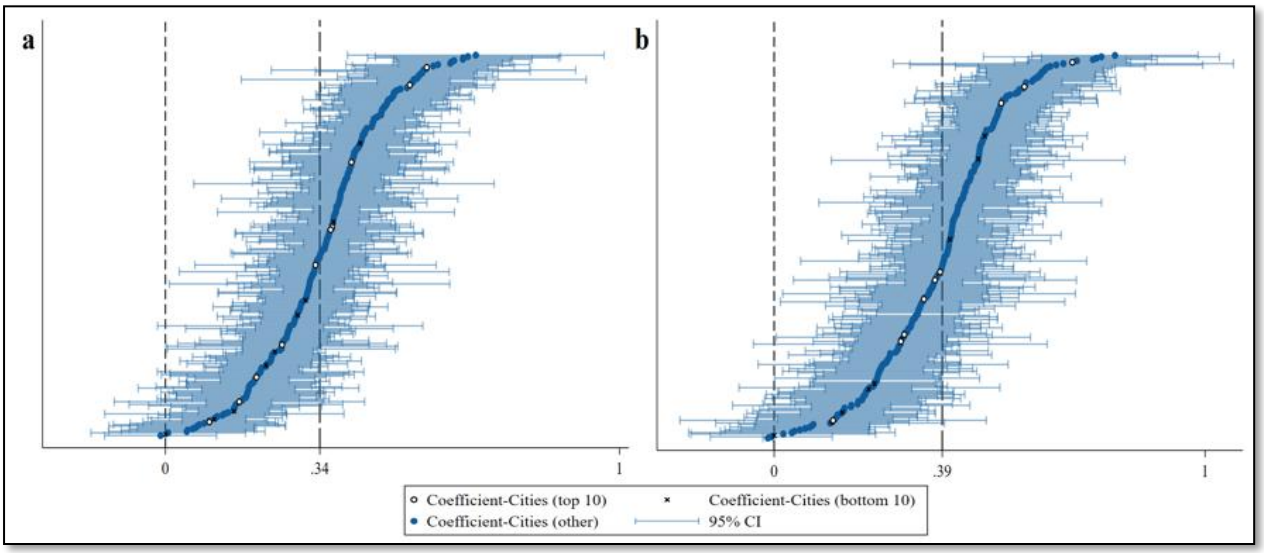


Figure A4: Spillover Effects of Prior Presence of Non-Academic PST in Adjacent School Districts[8]

Subsample - School Districts Bordering other Districts

Figure A5 presents city-specific spillover estimates for districts with at least one adjacent school district. Figure A5a shows estimates for all school districts, regardless of their 2015 PST status. Figure. A5b focuses exclusively on districts without local PST institutions in 2015. The coefficients are displayed in ascending order.

Figure A5a shows a range of coefficients from –0.041 to 0.830 (mean = 0.368). A comparable distribution is evident in Figure A5b for districts that initially lacked PST institutions, where coefficients span from –0.096 to 0.830 (mean = 0.394). The spillover effects for academic PST in the 10 cities that experienced the most pronounced changes in the share of one-child families (2015–2018) are consistent with the relationship depicted in Figure 5 of the main text.

---

specific mean spillover estimates for districts without a local PST institution in 2015. In both panels, cities are grouped along the x-axis into three categories based on the change in the share of one-child families: less than −0.1, between −0.1 and 0.1, and greater than 0.1.

[8] Figure A4a depicts city-specific estimates of spillover effects across all districts within each city. N=282. Figure A4b depicts city-specific estimates of spillover effects for districts without local PST institution in 2015 within each city; N = 282. In Figure A4a and A4b, the estimates are plotted with 95% CIs.

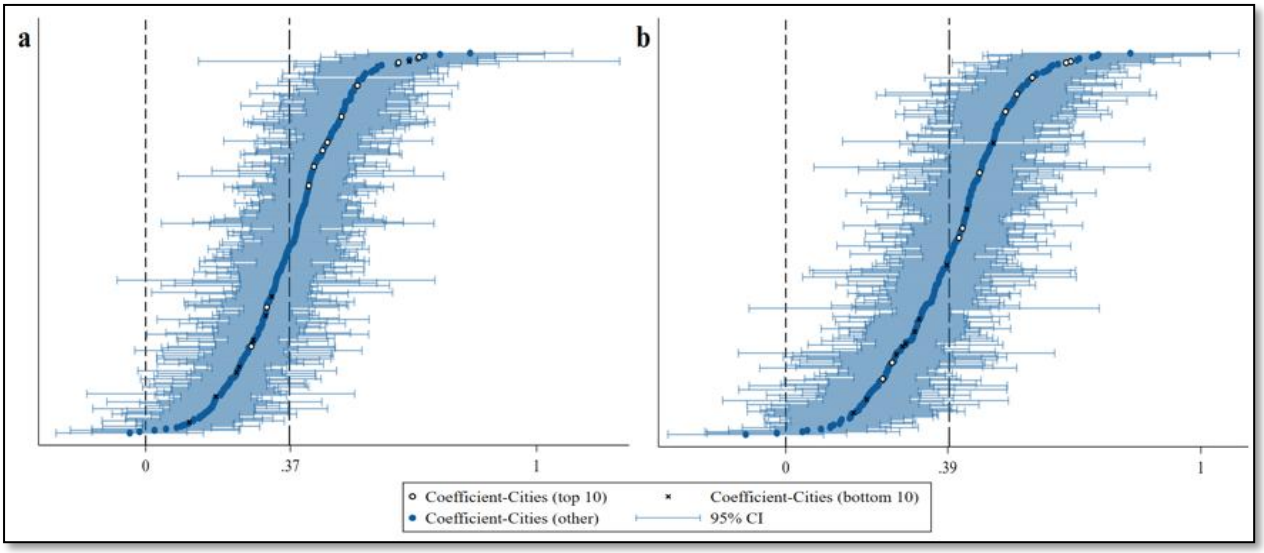


Figure A5: Spillover Effects of Prior Presence of PST in Adjacent School Districts Restricted to Districts with at Least One Neighbor[9]

A8. Declarations

Data Availability

The replication package, including processed datasets and analysis code (Stata do-files and Python NLP notebook), is publicly available at the Open Science Framework:

https://osf.io/9ktru/?view_only=3d0ff522d6374394aead79865a3c1d72.

The raw Point of Interest (POI) data were retrieved from the Amap open platform via its Application Programming Interface (API) (https://lbs.amap.com/api/webservice/guide/api-advanced/search); records have been cleaned and de-identified prior to inclusion in the replication package. The demographic data on one-child family shares were derived from the NBS–Peking University Research Data Center (https://nprdc.pku.edu.cn/info/1011/1274.htm); small random noise has been added to values for data protection.

Author Contributions

J.H., Z.C. and N.Z. contributed equally to data collection, NLP classification, and data analysis. J.H., Z.C., N.Z., J.S. contributed to conceptualization and interpretation of results. J.H. and J.S. contributed to writing and reviewing of the manuscript.

Competing Interests

The authors declare no competing interests.

Ethical Approval

This article does not contain any studies with human participants performed by any of the authors.

[9] Figure A5a depicts city-specific estimates of spillover effects across all districts within each city; $N = 282$. Figure A5b depicts city-specific estimates of spillover effects for districts without local PST institution in 2015 within each city; $N = 282$. In Figure A5a and Figure A5b, the estimates are plotted with 95% CIs.

Informed Consent

This article does not contain any studies with human participants performed by any of the authors.