Off-the-Job Learning in Cities*

Atsushi Yamagishi†

February 24, 2025

Abstract

Cities are characterized by high skill levels and workers improve their skills through learning. This paper presents novel stylized facts on learning behavior and cities. I use Japanese survey data that provide distinctively rich first-hand information about the frequency, purpose, and subject of off-the-job learning. First, people engage in learning more frequently in denser cities. Second, people in denser cities are more likely to learn to gain new employment or cultivate themselves. Third, while cities tend to facilitate learning of skills required in skill-intensive sectors, what people learn depends on the local demand for skills.

Keywords: Off-the-job learning, Agglomeration economies, Comparative advantage, Human capital accumulation, Cities

JEL classification: F11, J31, J62, R11, R12,

*I am grateful to Steve Redding and Leah Boustan for their guidance. I am also thankful to Jorge De la Roca (editor), two anonymous referees, Ikuto Aiba, Ellora Derenoncourt, Shota Fujishima, Keisuke Kawata, Mariko Nakagawa, Kentaro Nakajima, Hayato Nishi, Karthik Sastry, Yasuhiro Sato, Kohei Takeda, Xuanli Zhu, and participants of seminars and conferences for their helpful comments. I acknowledge permission to access the microdata of the Japanese time-use survey (*shakai seikatsu kihon chosa*), granted by the Ministry of Internal Affairs and Communications. I thank Yasuhiro Sato for his help in obtaining said permission. I acknowledge financial support from the Funai Foundation for Information Technology. All errors are mine.

†Institute of Economic Research, Hitotsubashi University; Center for Real Estate Innovation, the University of Tokyo; Global Japan Lab, Princeton University. E-mail: a.yamagishi@r.hit-u.ac.jp.

1 Introduction

Cities are characterized by workers' high skill levels (Behrens, Duranton and Robert-Nicoud 2014; Davis and Dingel 2020). While cities could attain higher skills through the sorting of high-skilled workers, cities could also attain them by facilitating the learning of each worker. Indeed, faster skill improvement through learning in cities has been identified as a key urban agglomeration force that justifies higher wages in cities (Duranton and Puga 2004). The idea that people learn faster in cities has featured in various theoretical models.[1] Empirically, faster wage growth in cities is often interpreted as empirical evidence consistent with faster learning in cities.[2] However, while important, indirect inference of the importance of learning from wage dynamics remains somewhat inconclusive because alternative mechanisms, such as richer job availability and better labor market matching (e.g., Dauth, Findeisen, Moretti and Suedekum 2022; Eckert et al. 2022; Papageorgiou 2022), can generate a similar pattern. Furthermore, wage information does not reveal additional details about learning, such as frequency, purpose and subject. Direct observation of learning behavior would provide more detailed and transparent evidence regarding learning in cities, which sheds light on why cities are endowed with more skilled workers. However, this has been challenging because learning behavior is not typically observed in data.

This paper presents a set of novel stylized facts on the relationship between learning behavior and cities, using a distinctive Japanese survey that collects detailed first-hand information on off-the-job learning.[3] The survey questions are about investments in human capital by time and effort spent on off-the-job training and education, which I call "off-the-job learning" in this paper. A complementary approach for learning is the "cognitivist" approach by Bacolod, Blum, Rangel and Strange (2023), who highlight the learning capacity of individuals; the ability to solve novel problems using reasoning. In contrast, I focus on providing novel evidence about how workers invest in human capital even after entry into the labor force, and how this investment pattern varies with urbanity.

The richness of the data allows me to answer the following questions: (i) Do people engage in off-the-job learning more frequently in denser cities? (ii) Does learning purpose differ in denser cities? (iii) What skills are people in denser cities more likely to learn? Due to the lack of panel data and a natural experiment, I focus on presenting new descriptive stylized facts and do not claim causality in the main part of this paper.[4]

The answers to the above three questions on learning in cities are as follows. First, people spend more time on off-the-job learning in denser cities: A 1% increase in population density is associated with a 0.1% increase in the number of days spent learning. This suggests that the

[1] See, for example, Glaeser (1999), Davis and Dingel (2019), and De la Roca, Ottaviano and Puga (2023).

[2] See, for example, Glaeser and Maré (2001), Baum-Snow and Pavan (2012), De la Roca and Puga (2017), and Eckert, Hejlesen and Walsh (2022).

[3] A distinct strand of literature in labor economics (e.g., Lynch 1992; Gaulke 2021) has analyzed off-the-job training, but its relationship with urbanity has not been investigated. I also investigate new aspects of learning, such as frequency and purpose.

[4] See Section 4 on the issue of causality.

faster wage growth in cities might be partially attributed to more efforts of the workers. From the productivity viewpoint, more time spent on learning behavior can be a new source of the widely-documented productivity advantages of denser cities (Rosenthal and Strange 2004). Moreover, more learning in cities could bolster the endowment of skilled labor, which contributes to meeting a high skill demand in cities documented by various empirical studies (e.g., Atalay, Sotelo and Tannenbaum 2024). More learning behavior in denser cities contributes to enhancing the skill endowment of cities and specialization of cities in skill-intensive sectors, above and beyond the sorting of high-skilled workers (e.g., Davis and Dingel 2020).

Second, people in denser cities are more likely to engage in learning behavior to obtain a new job or cultivate themselves. Learning for new jobs is consistent with the evidence that more job opportunities are available in cities (e.g., Papageorgiou 2022) because the presence of more attractive job offerings in larger cities would provide a stronger incentive for workers to acquire the necessary skills to take them.[5] This also implies that matching with better-paying jobs in larger cities (Eckert et al. 2022) might be partially driven by workers' voluntary efforts to improve matching quality by meeting skill demands, which is also consistent with empirical evidence that job ads in larger cities are more demanding in skill requirements and more efforts would be required for meeting them (Atalay et al. 2024). This highlights the potential complementarity between matching and learning in explaining urban agglomeration economies, two of the three key agglomeration forces in Duranton and Puga (2004). I also find that people are less likely to learn skills to use them at their current job, further highlighting the importance of new job opportunities. Learning for cultivation implies more cultural activities in denser cities, which is consistent with the agglomeration of cultural activities or creative industries (e.g., Mitchell 2019; Tao, Ho, Luo and Sheng 2019; Borowiecki and Dahl 2021). This may relate to cities' endogenous amenities in that agglomeration may improve neighborhood quality through cultural amenities and residents' cultural sophistication (Shapiro 2006; Diamond 2016).

Finally, regarding what people learn, people in denser cities are generally more likely to learn any skills related to skill-intensive sectors, such as foreign language skills and computer skills. This implies that off-the-job learning may indeed reinforce the comparative advantage of cities in skill-intensive sectors. In stark contrast, elderly care skills are more likely to be learned in *less* dense cities. This is consistent with the rapid population aging in rural Japan because the high demand for elderly care skills could dominate and overturn the general tendency of more frequent learning behavior in denser cities. Thus, people seem to respond to local demand conditions for skills, where cities face higher demand for high-skilled white-collar jobs while rural areas may face higher demand for jobs that require more physical skills such as elderly care.[6]

Although my main purpose is to present descriptive stylized facts, I assess how well the above-mentioned results capture the causal effect of population density on learning bahavior. Available results suggest that the causal interpretation might hold. First, I address the endogeneity of

[5]For evidence that different jobs require different skillsets, see Autor, Levy and Murnane (2003) and Ikenaga and Kambayashi (2016).

[6]Consistent with this, Atalay et al. (2024) find that jobs in larger cities tend to require cognitive and social skills, while those in smaller cities tend to require physical skills.

population density by using a long lag, following Ciccone and Hall (1996). Second, I focus on non-movers as in Charlot and Duranton (2004) to get a sense on the impact of endogenous sorting. Third, I use the method of Oster (2019) to provide a lower-bound estimate of the causal effect. Finally, I consider additional control variables, occupation and industry concentration, to further mitigate the omitted variable bias. All of these suggest that the potential endogeneity does not significantly alter my results.

I then explore how the relationship between learning behavior and population density varies by individual characteristics. I first conduct subsample analyses with respect to education categories, working status, gender, student status, and elderly status. I then explore the effect heterogeneity with respect to age and the skill level of occupation. While I find that the qualitative results are often similar across different subgroups, I point out some interesting heterogeneity. For example, the association between learning frequency and population density is stronger for non-college graduates, which might suggest the "catching-up" incentive in human capital investment after entering the labor force.

Finally, I quantitatively assess how much urban earnings premium can be explained by the learning channel. In my data, I estimate that 1% increase in population density is associated with around 0.04% increase in earning. This urban earnings premium is consistent with the meta study by Ahlfeldt and Pietrostefani (2019). I then assess how much off-the-job learning could explain it, finding that 2.2% of the urban earnings premium might be attributed to off-the-job learning. Note, however, that this could be a lower-bound of the importance of the learning channel because my cross-section data ignore earnings gain that materialize in the future.

An important qualification is that my data do not contain information on on-the-job learning.[7] Notably, Charlot and Duranton (2004, 2006) employ distinctive cross-section data from France and analyzes workplace communication, which is an important medium of on-the-job learning through knowledge spillovers, and how it varies with city size. While my approach is analogous in using survey data to directly analyze how learning depends on urbanity, this paper newly introduces survey data on off-the-job learning with rich information about frequency, purpose, subject and learning method. Such detailed stylized facts on off-the-job learning are important in themselves, and might also be suggestive for better understanding on-the-job learning since there has been no data about on-the-job learning that are equally rich and detailed. In particular, since off-the-job learning analyzed in this research requires time investment (see Section 2), my results might be especially informative for on-the-job learning that requires active learning costs such as time. For instance, if a higher return for skills in denser cities drives people to more frequently engage in costly off-the-job learning, the same incentive would also promote costly on-the-job learning.[8]

This paper is organized as follows. Section 2 describes my data and empirical strategy. Section 3 presents my main descriptive results. Section 4 assesses the causality of my descriptive

[7]Since I investigate learning as an act to improve one's skills, another type of learning that is not analyzed in this research is learning about uncertainty in one's own skills or job matching quality (e.g., Papageorgiou 2014; 2022).

[8]Sandvik, Saouma, Seegert and Stanton (2020) suggest that actively incurring learning costs, rather than learning passively, is important for gaining more from knowledge spillovers at the workplace.

evidence. Section 5 investigates the heterogeneity of the association of learning behavior and population density. Section 6 assess the quantitative importance of learning channel in explaing urban earnings premium. Section 7 concludes.

2 Data and empirical strategy

Data. I use the Japanese time-use survey (JTUS, *shakai seikatsu kihon chosa*) to obtain data about learning behavior. The JTUS is designed as a time-use survey, analogous to the American Time-Use Survey (ATUS), administered by the Statistical Bureau of Japan, the Ministry of Internal Affairs and Communications.[9] The JTUS provides repeated cross-sectional data that have been available every five years since 1976. About 90,000 households are sampled in each wave of the survey and all members of the sampled households answer the questionnaire. I use the individual-level data, which consist of about 200,000 individuals in each wave. I focus mainly on the 2016 data.[10] The JTUS is designed to be nationally representative and maintains a very high response rate, 95% for the JTUS 2016, because a response is required by law.[11]

While the JTUS has a similar time diary question as the ATUS, I focus on the question about off-the-job learning behavior in the past year, as shown in Figure 1, for which the ATUS does not have a counterpart.[12] The question instructs "[p]lease indicate the item you aimed to enhance your knowledge or level of culture, or to use for your current work..., *excluding* those activities conducted as a job or schoolwork" and "[e]xclude those activities directly related to regular courses in school, or employee training courses."[13] These instructions imply that on-the-job learning and learning that occurs passively (e.g., just being surrounded by smart people unconsciously enhances one's skills) are excluded. Therefore, in this paper, I focus on off-the-job learning that is undertaken actively with the intention to improve one's knowledge or cultural sophistication.[14]

The question pertains to eight skills: English language, other foreign languages, computers etc., commerce or business, elderly care, home economics or housework, humanities/social or natural sciences, and art and culture. For each skill, three subsequent questions are posed. First, the questionnaire asks, in nine categories, how many days in a given year the respondent spent learning a particular skill. The total number of days spent learning is obtained by summing up

[9]The use of time-use survey has been rare in urban economics literature. Recent notable exceptions include Murphy (2018) and Su (2022).

[10]I use the 1986 JTUS data for analyzing non-movers (see Section 4). I have also confirmed using the 2006 JTUS that my results are not specific to 2016.

[11]The response rate is taken from https://www.soumu.go.jp/main_content/000617655.pdf (in Japanese, last accessed on August 27, 2022).

[12]I also use the JTUS time diary question for supplementary analyses (see footnote 15 and Appendix B). See Kuroda (2010) and Lee, Kawaguchi and Hamermesh (2012) for more details on the time diary question of the JTUS.

[13]Italics added by the author. Some wording is the author's and differs from the official English translation of the original Japanese questionnaire in Figure 1 to better capture the original sentences.

[14]As mentioned in the introduction, this focus on human capital investment is different from the complementary approach of Bacolod et al. (2023), who focuses on learning capacity, the ability to solve novel problems using reasoning.

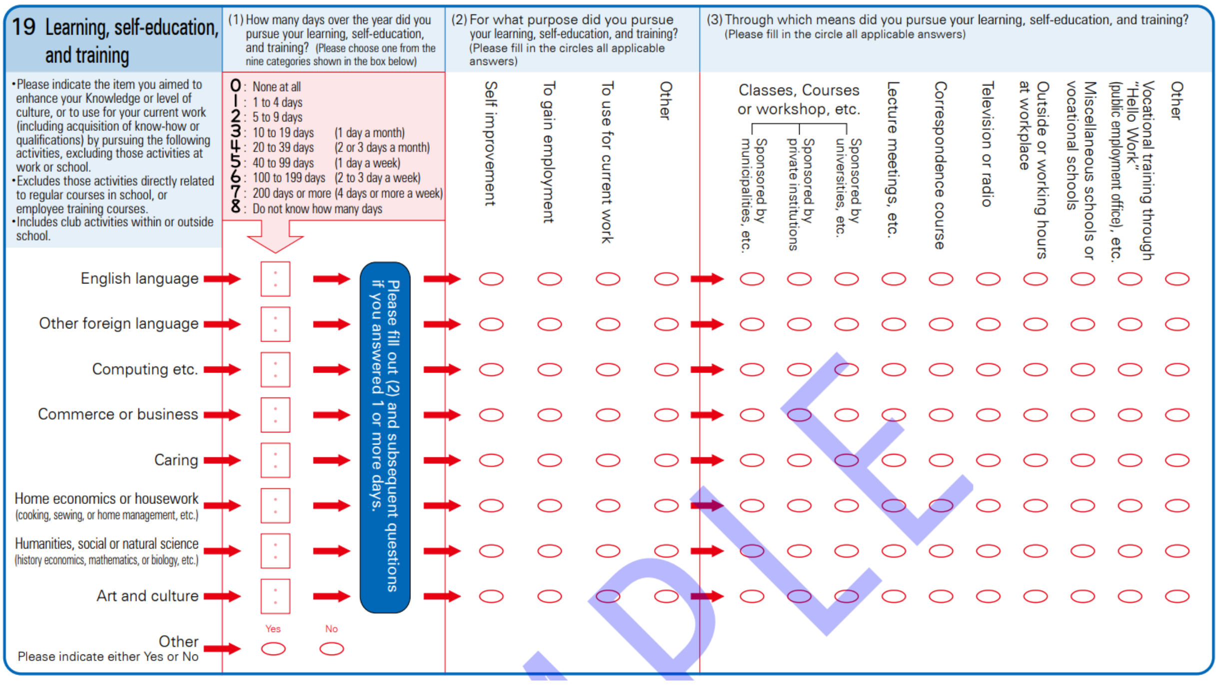


Figure 1: Questions on the off-the-job learning behavior (JTUS, 2016)

Note: The figure is an extract from the English translation of the questionnaire of the JTUS 2016, available at https://www.stat.go.jp/english/data/shakai/index.html (last accessed on August 17, 2022). The original questionnaire in Japanese can be found at https://www.stat.go.jp/data/shakai/2016/index.html (In Japanese. Last accessed on August 17, 2022).

the median value of each category across all skills.[15] Second, the questionnaire asks the purpose of learning the skill: for cultivating oneself, for gaining employment, for use on one's current job, or for other purposes (multiple choices are allowed).[16] Finally, the questionnaire asks how the respondent learned the skill(s) and lists the following options: classes provided by the public sector, classes provided by the private sector, classes provided by a university, lecture meetings, correspondence courses, television or radio, workplaces outside of working hours, vocational schools, training provided by the public employment office ("Hello Work"), and other means (multiple choices are allowed). In addition to the eight aforementioned skills, the survey also asks whether the respondent has learned anything that is not included among the eight skills. Overall, the questions allow me to investigate how often people engage in off-the-job learning activities, what they learn, for what purpose they engage in learning, and how they learn.

The survey also collects a standard set of socioeconomic characteristics for each respondent. Motivated by the urban wage premium literature (e.g., Glaeser and Maré 2001; Charlot and Duranton 2004; De la Roca and Puga 2017), I control for sex, age and its square, marital status, and educational attainment throughout this research. Since the survey question is designed for

[15]This might cause a double-counting issue if an individual learns skill A and skill B on the same day. To address this, I also apply an alternative measure of learning frequency taken from the time-diary question of the JTUS, which is the number of minutes spent on off-the-job learning in a given day. I find that people learn more in denser cities, both on the intensive and extensive margins.

[16]The official English translation shown in Figure 1 states "self improvement," but I believe that "for one's own cultivation" better captures the original Japanese sentence (*"jibun no kyouyou wo takameru tame"*).

analyzing the learning patterns of workers, I focus on people between the ages of 25 and 59 years to focus on the working-age population. The data also include an individual's occupation if they are working, which I control for in some specifications.[17] I also observe the annual income, which I use as an outcome variable in Section 6.

I obtain municipal population data from the 2015 Population Census, which is the closest census year to 2016.[18] I also take the habitable area of each municipality from the Municipal Area Statistics of Japan. I define the population density of each municipality by dividing the total population by the habitable area. As of 2016, there were 1,741 municipalities in Japan, and 1,350 municipalities are covered in my final JTUS 2016 dataset.[19]

To graphically illustrate the relationship between learning outcomes and population density in the raw data, Figure 2 plots key learning behaviors analyzed in Section 3 against municipal population density. It shows that they have apparent correlations that are not driven by a few outliers. The scatterplots for other learning outcomes are shown in Figure A.1.

Tables A.1 and A.2 present summary statistics. In Table A.2, individual characteristics are shown separately for learners (i.e., those engaging in some learning activities in a year) and non-learners (i.e., those not engaging in any learning activities in a year). Learners and non-learners are similar in age, gender, and marital status. However, learners tend to have a college degree or a higher income.

Empirical strategy. I analyze how various outcome variables related to off-the-job learning correlate with population density. The outcome of individual i living in municipality j is denoted by y_{ij}. I estimate the following linear model using ordinary least squares (OLS):

$$y_{ij} = \beta \ln PopDens_j + \gamma X_i + \epsilon_{ij}. \tag{1}$$

where $\ln PopDens_j$ is the log population density of municipality j, X_i represents the characteristics of individual i that I control for, and ϵ_{ij} is the error term.[20]

In equation (1), we cannot use as y_{ij} as the log of the number of days that individual i in municipality j engages in learning. In this case, I also estimate the following constant-elasticity regression equation:

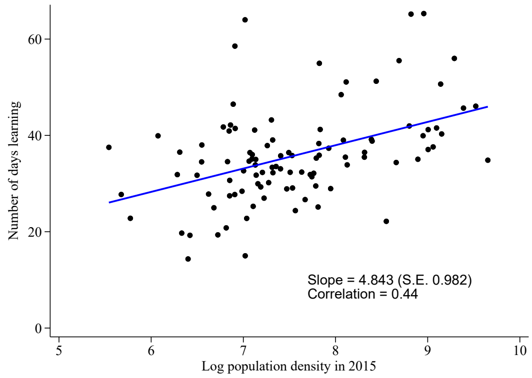
$$y_{ij} = \exp(\beta \ln PopDens_j + \gamma X_i)\eta_{ij}, \tag{2}$$

[17] I do not control for occupation in the main analysis because, as discussed later in this paper, learning for gaining new employment is an important motivation for learning but I cannot observe occupation for unemployed workers. I use the major categories (*dai bunrui*) of occupation classification. Unfortunately, the JTUS data do not record the industry of each worker.

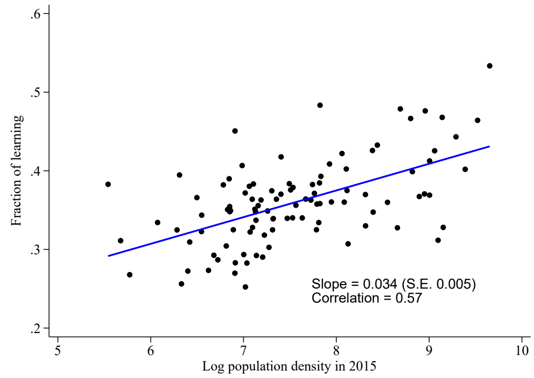
[18] The data for population density are downloaded from https://www5.cao.go.jp/keizai-shimon/kaigi/special/reform/mieruka/db_top/index.html (In Japanese. Last accessed on June 30, 2022).

[19] When analyzing the JTUS data prior to 2016 for robustness checks, I use the municipal boundaries as of 2015, while I use the municipal code converter, provided by Kondo (2022), to deal with municipal mergers.

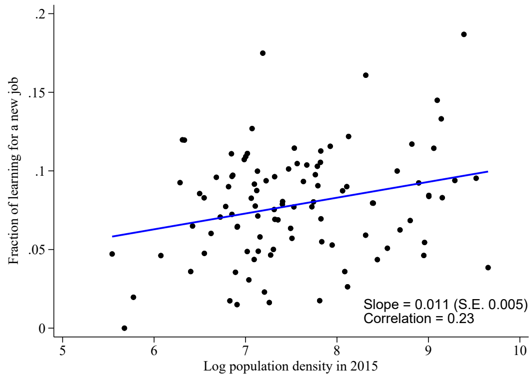
[20] While I use municipal population density to measure the density of economic activities, using the metropolitan-area level density defined by Kanemoto and Tokuoka (2002) does not change my conclusion.



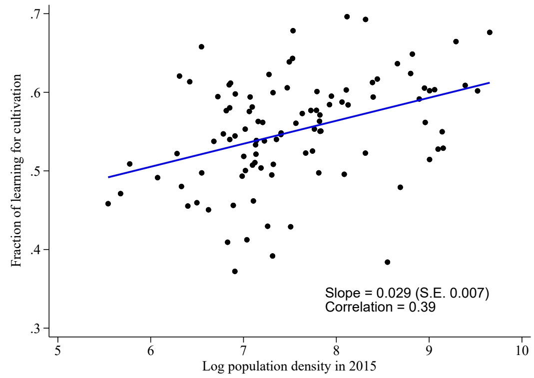
(a) Number of days engaging in learning



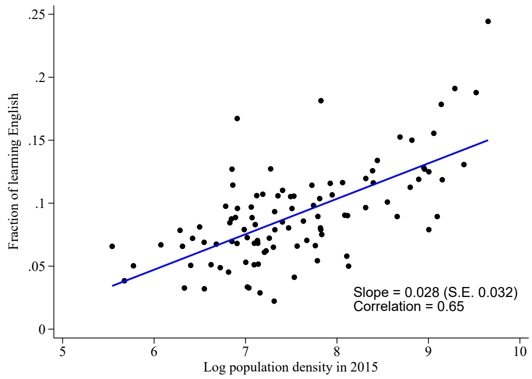
(b) Fraction of learning something



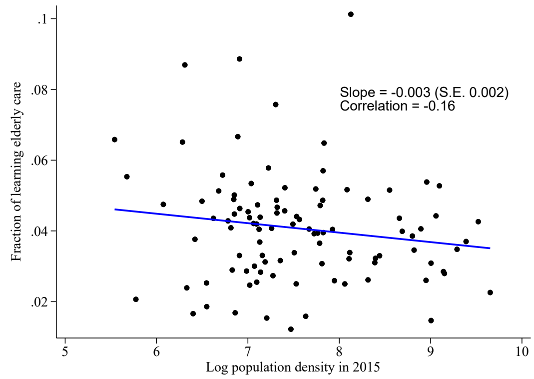
(c) Fraction of learning for a new job



(d) Fraction of learning for cultivation



(e) Fraction of learning English



(f) Fraction of learning elderly care

Figure 2: Selected learning behavior and population density

Note: The figures plot the municipal average of selected learning outcomes against its log population density in 2015 (black dots). The blue line is the regression line. The fractions for a purpose of learning is calculated among those who learned something. To credibly calculate municipal averages and to comply with the data privacy concern by the data provider, I focus on municipalities with more than 150 individual observations (101 municipalities) in these figures. Using alternative threshold numbers of minimum observations produces similar figures. Similar plots for the remaining learning outcomes are found in Figure A.1.

where y_{ij} is the number of days spent learning. β, the coefficient of interest, is interpreted as the elasticity of the outcome y_{ij} with respect to population density. η_{ij} is a multiplicative error term with. Importantly, y_{ij} can be zero, unlike in the standard linear model (1) with a logarithmic outcome variable. Allowing y_{ij} to take a value of zero may be important in this context because only around a third of my sample engages in off-the-job learning. I follow Santos Silva and Tenreyro (2006) and estimate (2) using the Poisson pseudo-maximum-likelihood (PPML) estimator.[21]

I interpret β in equations (1) and (2) as descriptive. This is because, unlike studies in the literature on the urban wage premium (e.g., Glaeser and Maré 2001; De la Roca and Puga 2017), I cannot control for individual fixed effects to address sorting bias based on unobserved individual characteristics (e.g., intrinsic motivation) due to the cross-sectional nature of the JTUS.[22] Moreover, there is no natural experiment that randomly affects population density in this context. Given these limitations, I focus on providing descriptive evidence as my main results (Section 3). The issue of causality is relegated to Section 4,

Throughout this research, I ensure the national representativeness of the data by weighting observations according to the sampling probability provided in the dataset. I cluster standard errors at the municipal level.

3 Main descriptive results

3.1 Learning frequency

I first analyze how much more frequently people engage in off-the-job learning in denser cities. The estimation results in Column 1 of Table 1 show that in specification (2), a 1% increase in population density is associated with a 0.1% increase in the number of days spent learning. Since the standard deviation of the log population density is 1.209 (Table A.2), a one standard deviation increase in population density is associated with around 12% increase in learning frequency.[23]

This positive effect is driven by both intensive and extensive margins. Column 2 focuses on the intensive margin by using the log of the number of days in the linear regression (1), and Column 3 focuses on the extensive margin by using a dummy variable indicating whether whether an individual engages in any learning. I find positive and statistically significant coefficients in both columns. Studies have documented faster wage growth in cities (e.g., Glaeser and Maré 2001; Baum-Snow and Pavan 2012; De la Roca and Puga 2017; Eckert et al. 2022). More

[21]As an alternative method, I apply the inverse hyperbolic sine transformation $\ln(y_{ij} + \sqrt{y_{ij}^2 + 1})$, where y_{ij} is the number of days that the individual spends in learning and this can be zero (Bellemare and Wichman 2020). By taking this as the outcome variable and estimating the model (1) using OLS, the estimated elasticity with respect to population density is 0.109, which differs little from 0.102 obtained by the PPML estimation in Table 1.

[22]Note, however, that even individual fixed effects cannot resolve the endogeneity due to sorting if the sorting is based on the unobserved individual-specific slope of population density (i.e., sorting based on individual-specific β).

[23]For a more specific example, in 2016, Osaka City had a log population density of around 9.4, while Hiroshima City had a log population density of around 7.2. This implies that people in Osaka spend about 0.1×2.2 =22% more days learning than those in Hiroshima.

	Number of days learning (1)	Log number of days learning (2)	Engaged in learning (3)	
ln population density	0.102[a]	0.053[a]	0.021[a]	
	(0.017)	(0.016)	(0.003)	
Estimation method	PPML	OLS	OLS	
Oster's bound	NA	0.047	0.014	
N	78727	23513	78727	
R^2	NA	0.014	0.069	
	Learning for new employment (4)	Learning for current job (5)	Learning for cultivation (6)	Learning for other objectives (7)
ln population density	0.008[b]	-0.009[b]	0.026[a]	0.004
	(0.004)	(0.004)	(0.004)	(0.005)
Estimation method	OLS	OLS	OLS	OLS
Oster's bound	0.009	-0.013	0.023	0.006
N	24905	24905	24475	24905
R^2	0.016	0.070	0.032	0.025

Cluster-robust standard errors in parentheses

^{c} $p < 0.1$, ^{b} $p < 0.05$, ^{a} $p < 0.01$

Table 1: Learning frequency and purposes

Note: In Column 1, I estimate by the Pseudo Poisson Maximum Likelihood (PPML, Santos Silva and Tenreyro 2006) the constant-elasticity regression model (2) relating number of days spent for learning to the municipal population density. Column 2 takes as the outcome variable the log number of days spent for learning in the linear regression model (1). The remaining columns use the linear probability model as the outcome variable is a dummy. Column 6 has slightly fewer observations than columns 4, 5, and 7 because some observations are coded as "unknown." In all specifications, I include as control variables sex, age and its square, marital status, and educational attainment. I calculate Oster's (2019) beta for each regression. Based on a recommendation of Oster (2019), I set the maximal R^2 to 1.3 times the R^2 of each regression. I also assume that observed and unobserved factors are equally related to population density (i.e., the coefficient of proportionality $\delta = 1$ in Oster's notation). Due to its availability, I do not report it when using the PPML. I cluster standard errors at the municipal level.

frequent learning in denser cities is consistent with this finding.[24]

Longer hours dedicated to learning may provide a new microfoundation for urban agglomeration economies in terms of productivity. Duranton and Puga (2004) point out that learning, matching, and sharing constitute three important classes of mechanisms contributing to higher urban productivity. While larger cities are known to demand higher skill levels of workers (e.g., Rosenthal and Strange 2004; Atalay et al. 2024), studies have emphasized the efficiency of learning in cities for a given amount of learning time, especially through knowledge spillovers (e.g., Marshall 1890; Duranton and Puga 2004; Charlot and Duranton 2004; Davis and Dingel 2019). On the other hand, my findings indicates a novel simple microfoundation for higher productivity in denser cities based on learning: People in denser cities invest more time in improving their

[24] I also investigate the heterogeneous impact of density across different socioeconomic groups (see Section 5).

productivity.[25] Taken together, it is likely that both the quality and quantity of learning are greater in denser cities, both of which contribute to their higher productivity.[26]

More frequent learning behavior in cities also provides a new microfoundation for the comparative advantage of cities in skill-intensive sectors. More learning in cities may bolster the supply of skilled labor beyond the mere sorting of high-skilled workers into urban areas. Davis and Dingel (2020) show that cities attract high-skilled workers, which influences trade patterns by leading cities to specialize in skill-intensive sectors. Therefore, my findings suggest that more learning in cities could further reinforce their comparative advantage in skill-intensive sectors and drive greater specialization in these sectors. Consistent with this, Appendix F demonstrates that high-skill jobs are more prevalent in denser cities within the Japanese context.

3.2 Purposes of learning

I now analyze the purpose of off-the-job learning behavior. As an outcome, I use a dummy variable indicating whether a person engages in learning for a given purpose. I consider three purposes, namely, (i) to obtain a new job, (ii) to utilize the skill at their current job, and (iii) for personal cultivation. For completeness, I also report results for learning motivated by other reasons. To focus on the association between learning purposes and population density while controlling for the effect of learning frequency, the sample in this analysis includes only individuals who engaged in some learning. This results in a smaller sample size than in Section 3.1.

Columns 4–7 of Table 1 report the results. Column 4 shows that people in denser cities are more likely to learn with the goal of obtaining a new job. This is consistent with the idea that people are more likely to find new job opportunities (Di Addario 2011; Papageorgiou 2022) in denser cities and thus have a greater incentive to meet the skills requirements of the new job opportunities (c.f., Autor et al. 2003; Ikenaga and Kambayashi 2016). Indeed, Eckert et al. (2022) recently found that gradual matching with more skill-intensive occupations and industries can explain a significant part of the faster wage growth observed in a denser city. My findings suggest that such a transition might be partially facilitated by workers' voluntary learning to improve their skills, which is also consistent with Atalay et al. (2024), who show that jobs in larger cities tend to be more skill-intensive. This highlights the potential complementarity between matching and learning in explaining urban agglomeration economies.[27] Column 5 shows that, in stark contrast to Column 4, people are *less* likely to learn specifically to apply skills in their current jobs. This result implies that workers in denser cities are relatively more motivated by new job opportunities rather than improving productivity within their current job. In this sense, these

[25] This is consistent with the theoretical prediction of Davis and Dingel (2019).

[26] It is important to note that the welfare implications of agglomeration economies depend on whether learning quality or quantity improves in cities. Generally, the welfare gain depends on whether this effect represents a pure positive externality, or if comes at the expense of other resources. Here, the higher productivity induced by more frequent learning is offset by the cost of people's time investment. Therefore, misattributing such productivity gains resulting from more learning to other sources of agglomeration economies, such as more efficient learning through knowledge spillovers, could lead to substantially different welfare implications (c.f., Miyauchi 2024).

[27] Learning and matching are two of the three key urban agglomeration forces identified by Duranton and Puga (2004).

findings further reinforce the view that greater job availability in urban labor markets incentivizes people to acquire relevant skills.

In addition to labor market concerns, in Column 6, I find that people in denser cities also learn to cultivate themselves.[28] More learning for cultivation in denser cities is consistent with the agglomeration of cultural activities and creative industries (e.g., Mitchell 2019; Tao et al. 2019; Borowiecki and Dahl 2021). This may also relate to studies on endogenous urban amenities (e.g., Shapiro 2006; Diamond 2016). My findings highlight that urban agglomeration may enhance cultural activities, potentially fostering more cultural amenities or a greater degree of cultural engagement among residents.[29] In line with this, I find that the proportion of cultural jobs increases with population density (see Appendix F). Incorporating this channel into theories of endogenous amenities might be important.

3.3 What people learn

I investigate the types of skills people learn. For each skill, I create a dummy variable indicating whether a person spent some time learning it and use it as the dependent variable.

Table 2 presents the results on the types of skills people acquire. Almost all skills are more likely to be learned in denser cities, which is intuitive given that people engage in learning more frequently. This implies that cities indeed accumulate skills required for skill-intensive sectors, such as foreign languages and computer-related skills, thereby bolstering their comparative advantage in such industries. This is consistent with prior literature demonstrating that larger cities tend to host occupations with substantial cognitive and interactive content rather than those requiring primarily physical or manual labor (e.g., Bacolod, Blum and Strange 2009; Atalay et al. 2024).

There is, however, an exception: elderly care skills are more likely to be learned in less densely populated cities.[30] I interpret this as evidence that the skills people acquire are responsive to skill demand in the local labor market. The demand for elderly care is increasing in rural Japan due to an aging population, making it a relatively attractive employment option compared to other sectors (Hanaoka 2015).[31] Presumably, strong local demand for elderly care skills outweighs the general tendency for skill acquisition to be more prevalent in denser cities.[32] Consistent with this,

[28]A possible explanation is that the availability of cultural facilities, such as museums, encourages individuals to acquire related knowledge (e.g., art history) to enhance their enjoyment of cultural activities (Krupka 2009). In Section 5, I find no significant heterogeneity in learning for cultivation across various characteristics (e.g., age, gender, education), which is expected, as access to cultural facilities is broadly available across different socioeconomic characteristics.

[29]Consistent with this, Boualam (2014) suggests that, although statistically insignificant, culture might contribute to improvements in consumption amenities.

[30]Given that elderly care is likely to require manual skills more intensively, this is consistent with Atalay et al. (2024), who document the prevalence of manual jobs in smaller cities.

[31]Indeed, the proportion of care workers in the total workforce is higher in rural prefectures (https://www.sssc.or.jp/touroku/tourokusya.html, in Japanese. Last accessed on June 2, 2022). A possible explanation is that, due to price regulations in the elderly care sector, the relative wage rate in this sector is higher in rural areas than in urban areas (Hanaoka 2015).

[32]In Section 5, I find that the positive association between population density and learning elderly care skills is

	English	Other foreign language	Computers etc	Commerce and Business	Elderly care
	(1)	(2)	(3)	(4)	(5)
ln population density	0.020[a]	0.007[a]	0.010[a]	0.010[a]	-0.004[a]
	(0.002)	(0.001)	(0.002)	(0.002)	(0.001)
Estimation method	OLS	OLS	OLS	OLS	OLS
Oster's bound	0.015	0.006	0.007	0.008	-0.004
N	78727	78727	78727	78727	78727
R^2	0.080	0.016	0.027	0.039	0.013
	Housekeeping	Humanities and sciences	Arts and culture	Other	
	(6)	(7)	(8)	(9)	
ln population density	0.007[a]	0.004[b]	0.013[a]	0.000	
	(0.002)	(0.002)	(0.002)	(0.002)	
Estimation method	OLS	OLS	OLS	OLS	
Oster's bound	0.007	0.001	0.010	-0.001	
N	78727	78727	78727	78727	
R^2	0.036	0.050	0.037	0.001	

Cluster-robust standard errors in parentheses

[c] $p < 0.1$, [b] $p < 0.05$, [a] $p < 0.01$

Table 2: Learning by subject

Note: I use the linear probability model that regresses the dummy for learning a particular subject on the municipal population density. In all specifications, I include as control variables sex, age and its square, marital status, and educational attainment. I calculate Oster's (2019) beta for each regression. Based on a recommendation of Oster (2019), I set the maximal R^2 to 1.3 times the R^2 of each regression. I also assume that observed and unobserved factors are equally related to population density (i.e., the coefficient of proportionality $\delta = 1$ in Oster's notation). I cluster standard errors at the municipal level.

Section F confirms that rural areas indeed exhibit higher demand for elderly care workers than urban areas. Overall, the results in Table 2 highlight that skill acquisition patterns are responsive to local labor demand conditions.

Additional results In Appendix B, I provide additional results on (1) the methods through which people acquire skills and (2) the individuals with whom they engage in learning. In summary, I find that market-based education methods (e.g., private classes) play a more significant role in denser cities. Additionally, I find no correlation between urbanity and the likelihood of engaging in off-the-job learning with colleagues, classmates, or others.

4 Causality of urbanity on learning behavior

Although the previous section primarily aimed to descriptively establish stylized facts regarding the association between learning behavior and population density, I also assess the extent to which my estimates can be interpreted causally through several additional analyses. While not definitive, the following analyses suggest that the estimated β in Section 3 may approximate a causal relationship.

Endogeneity of population. Though not a problem for a descriptive analysis, identifying the causal impact of population density on learning behavior is impossible if some unobserved factors simultaneously affect population and learning. To address this, I follow Ciccone and Hall (1996) and use as an instrumental variable (IV) a long lag of the population density using the 1920 population census, the first modern population census in Japan. I find that, although somewhat noisier, the IV results are similar to the baseline OLS results. This is in line with studies on the urban wage premium that the endogeneity of city size has a limited impact on results (e.g., Ciccone and Hall 1996; De la Roca and Puga 2017). Details are in Appendix C.

Sorting and non-movers. I follow Charlot and Duranton (2004) and focus on non-movers to assess how endogenous sorting might affect my results. Intuitively, if the sorting by unobserved factors drives the observed correlation between learning behavior and density, then it might accompany geographical mobility because people have an incentive to reallocate to a location that best matches with their type. Note that endogenous sorting might not necessarily entail geographical mobility, and so focusing on non-movers might not fully eliminate the effect of endogenous sorting. For instance, those born in higher density areas might have better ability due to better school quality. That said, it might alleviate the correlation between residential location and certain unobserved characteristics (see Combes, Duranton and Gobillon 2011 for more discussions). Since the 2016 JTUS data do not contain questions about the previous

stronger for women. This is reasonable, given that the elderly care sector is a female-dominated industry and thus more relevant for women.

residence, I investigate the 1986 JTUS data containing this information.[33]

I investigate the heterogeneous association of population density with these three types of learning outcomes by analyzing the subsamples of nonmovers defined in two ways: (i) the person stayed in the same residence for more than four years or (ii) the prefecture in which a person went through compulsory education was the same as the current prefecture. I find little evidence that the relationship between learning behavior and population density is different among the sample of non-movers. This might suggest the relatively limited importance of endogenous sorting. See Appendix D for details.

Oster's bound. In Tables 1, 2, and B.1, I report Oster's (2019) beta. Oster's beta is a conservative estimate of β for causality, taking into account the presence of unobserved selection bias into denser cities. As such, Oster's bound is an approach to address endogenous sorting, which yields correlation between unobserved factors and urbanity. Following Oster (2019), I assume that observed and unobserved factors are equally related to population density. I also assume that the R-squared when controlling for both observed and unobserved variables equals 1.3 times the R-squared of the regression with observed control variables. I find that no statistically-significant result changes qualitatively, alleviating the concern that endogenous sorting might spuriously drive my results.

Additional control variables. I consider some additional control variables to mitigate the concern of omitted variables. First, to check whether occupation composition that varies with urbanity affects my results, I additionally control for the dummies of individual occupation categories. Tables A.3 and A.4 show that our main results in Tables 1 and 2 are robust to controlling for them.

Second, I consider the degree of industrial concentration in local labor markets, which is a key determinant of local labor market outcomes (Rosenthal and Strange 2004; Azar, Marinescu, Steinbaum and Taska 2020). To measure the degree of industry concentration, I construct the Herfindahl-Hirschman index (HHI) using the employment share of each major industry category (*sangyo dai bunrui*). I then include the HHI as an additional regressor to our main analysis (Tables 1 and 2). Tables A.5 and A.6 present the results. Overall, the estimates are similar to the main results.

5 Subsample analysis and treatment effect heterogeneity

This section first conducts subsample analysis by education category, working status, gender, student status, elderly status, age, and the skill level of an occupation. Table 3 provides a summary of the heterogeneity analysis in this section for the key learning variables, and the Appendix tables described in the table footnote provide further details.

[33]The 1986 JTUS asks frequency, subject, and method of learning in a different format than the 2016 JTUS. Unfortunately, the purpose of learning was not asked in the 1986 JTUS.

Subsample	Number of days learning	Log number of days learning	Engaged in learning	Learning for new employment	Learning English	Learning computer skills
Baseline	0.102	0.053	0.021	0.008	0.020	0.010
College graduates	0.077	0.071	0.025	0.001	0.027	0.008
Non-college graduates	0.124	0.043	0.019	0.013	0.017	0.010
Devoted to work	0.101	0.040	0.024	0.008	0.020	0.010
Not devoted to work	0.080	0.044	0.017	0.002	0.020	0.009
Male	0.143	0.098	0.024	0.015	0.020	0.013
Female	0.063	0.012	0.017	0.001	0.020	0.006
Students	0.025	0.033	0.016	-0.001	0.020	0.000
Elderly	0.067	0.046	0.011	0.002	0.006	0.006
Young	0.129	0.036	0.035	0.003	0.028	0.026
Not young	0.096	0.056	0.019	0.010	0.019	0.007
High-skill job	0.068	0.076	0.024	0.013	0.027	0.021
Not high-skill job	0.114	0.046	0.021	0.007	0.019	0.008

Table 3: Summary of heterogeneity analysis

Note: The table summarizes the heterogeneity about the coefficient of log population density on selected learning outcomes. The baseline result comes from Tables 1 and 2. The results on college graduates and non-college graduates come from Tables A.7–A.10. The results on working status (devoted or not devoted to work) come from Tables A.11–A.14. The results on the male and female samples come from Tables A.15–A.18. The results on students come from Tables A.19 and A.20. The results on the elderly come from Tables A.21 and A.22. The results on young and non-young workers come from Tables A.23 and A.24. The results on high-skill and non high-skill jobs come from Tables A.25 and A.26.

Estimation by education categories. Although the results for both college and non-college graduates are overall similar, there are a few important differences. First, in terms of learning purpose, job-related learning ("learning for new employment" and "learning for current job") is significant only for non-college graduates. This implies that off-the-job learning might be more important for workers who have not yet accumulated enough skills to take up a job requiring high skills. As a result, they might be trying to "catch up" in their human capital by investing more time and effort after entering the labor force. Second, learning an elderly care skill has a significant negative association with population density only for noncollege graduates. This is consistent with a lower share of workers with a college degree in the elderly care industry.[34]

Estimation by working status. Overall, results are similar for workers who are mainly devoted to working and those who are not (e.g., the unemployed and housewives). One notable difference is that learning for acquiring a new job is significant only for workers who are mainly devoted to working. This suggests that in denser cities, workers who already have a job are more motivated to acquire a new job. This is consistent with the importance of the job ladder in cities, consistent with Eckert et al. (2022).

Estimation by gender. While results are qualitatively similar for both men and women, there are two important differences. First, the positive association between learning frequency and population density is stronger for men. Indeed, Table 3 reveals that the elasticity of learning

[34]Source: https://www.mof.go.jp/pri/publication/research_paper_staff_report/staff23.pdf (in Japanese, last accessed on June 29, 2024).

frequency for men is more than twice as large as that for women. This implies that the incentive for learning in urban areas can be stronger for men, which might be natural given that women tend to spend fewer years in the labor force. Second, learning for a new job is significant only for men, which might suggest that the incentive to climb up a job ladder in urban areas is more important for men.

Students and the elderly. While the results for students and adult workers are often qualitatively similar, there are some important differences between students and adult workers. First, there is a much weaker association between learning frequency and population density for students. Second, students in denser areas do not learn more for new employment. Overall, these results suggest a weaker incentive for students to engage in off-the-job learning, suggesting the importance of labor market incentives that students may not take seriously.[35]

The results for the elderly (those over the age of 60) and adult workers are also often qualitatively similar. Similarly to the case of students, an important difference is that the association between learning frequency and population density is much weaker for the elderly than for adult workers. This again highlights the importance of labor market incentives for off-the-job learning behavior.

Heterogeneous effect by age. We may expect that young people are more responsive to population density in determining their learning behavior. Generally, young workers may respond more to learning incentives, which depend on urbanity, because young workers are expected to spend more time in the labor market. Moreover, Bacolod et al. (2023) show that learning capacity decays with age, implying that younger people have a higher return on engaging in learning behavior. To test such a heterogeneous effect by age, I add the interaction term between population density and a young dummy (i.e., a dummy variable indicating whether a person is 30 years old or younger).[36] I find qualitatively that the association between learning frequency and population density is stronger for young workers, but these terms are generally statistically insignificant. One statistically significant heterogeneity is that young workers are even more likely to learn computer skills in denser cities, which seems natural given the importance of computer skills in the job postings of dense cities (Atalay et al. 2024).

Heterogeneous effect by the skill level of occupation. To the main specification, I add the interaction term between population density and a high-skill occupation dummy (i.e., a dummy variable indicating whether the person engages in a high-skill occupation).[37] Overall, the in-

[35]Note that the survey questionnaire in Table 1 explicitly excludes learning at school. Results in Tables A.19 suggest that learning behavior out of school is less associated with population density for students.

[36] Here I employ an interaction term specification, rather than splitting the sample, to facilitate a test of statistical significance. This is important as we find relatively large differences in the coefficients in the point estimates (see Table 3). Subsample analysis has the advantage that it allows for arbitrary heterogeneity in the coefficients of control variables (c.f., Kawaguchi 2011). Using the subsample analysis leads to similar estimates (available upon request).

[37]'High-skill occupation" is defined as belonging to the major occupation classification "professional and engineering worker." See https://www.stat.go.jp/english/data/roudou/report/2012/pdf/ap06.pdf (last accessed on June

teraction term does not exhibit strong statistical significance. There are two notable exceptions. First, learning for the current job is even more active in denser cities if workers have a high-skill occupation, which seems natural given the high skill requirement of the current job. Second, for computer skills and elderly care skills, workers with a high-skill occupation are more responsive to population density. This is consistent with the prevalence of official licenses in these industries, which are often required for professional and managerial positions.[38]

6 Quantitative importance in accounting for urban earnings premium

In this section I examine the relationship between earnings and population density and how much of it can be explained by the learning channel I have highlighted so far. I first replicate the earnings premium in my JTUS data.[39] I regress the log annual earnings on the log population density and the same set of control variables as in Section 3. I find that a 1% increase in density is associated with a 0.039% increase in income. The magnitude of this estimate is close to 0.04% in the meta-analysis of the urban wage premium by Ahlfeldt and Pietrostefani (2019).

To quantitatively assess how much of such urban earnings premium can be explained by the learning channel, I follow Charlot and Duranton (2004) and compute how much of the effect of population density on individual earnings can be explained by off-the-job learning behavior. Specifically, I first regress log individual earnings on log population density, the inverse hyperbolic sine transformation of the number of days learning, and the control variables.[40] That is, I estimate the following equation:

$$\ln\,(\text{Earnings})_{ij} = \gamma_1 \ln PopDens_j + \gamma_2 \ln\left(n + \sqrt{n^2+1}\right)_{ij} + \gamma_E X_i + \epsilon_{ij}, \tag{3}$$

where n denotes the number of days learning. I next regress on the (inverse hyporbolic sine transformation of the) number of days learning on the log population density and the same control variables, and I denote by γ_3 the coefficient of the log population density. That is, I estimate the following equation:

$$\ln\left(n + \sqrt{n^2+1}\right)_{ij} = \gamma_3 \ln PopDens_j + \gamma_L X_i + \epsilon_{ij}. \tag{4}$$

Then, $\gamma_2\gamma_3/(\gamma_1+\gamma_2\gamma_3)$ gives the fraction of the effects of population density that is mediated

30, 2024). Note also that we use the interaction term specification for the same reason as footnote 36.

[38]See https://www.jil.go.jp/institute/reports/2010/documents/0121-1_02.pdf for the prevalence of official licenses in the elderly care and information industries (in Japanese, last accessed on June 30, 2024).

[39]Although it has already been shown that larger cities have higher income in Japan (e.g., Tabuchi and Yoshida 2000), my analysis would offer some new insights by using more recent individual-level microdata.

[40]Instead of the inverse hyperbolic sine transformation, the transformation of adding one and applying the log transformation hardly changes the conclusion. See the table footnote of Table E.1 for the list of control variables.

through the number of days learning. I estimate that $\gamma_1 = 0.038$ (S.E. 0.006), $\gamma_2 = 0.008$ (S.E. 0.003), and $\gamma_3 = 0.107$ (S.E. 0.015). I get $\gamma_2\gamma_3/(\gamma_1 + \gamma_2\gamma_3) \simeq 0.022$, suggesting that 2.2% of the population density effect is explained by the number of days learning.[41] Note, however, that this may be an underestimation of the importance of learning behavior on earnings, as my cross-sectional data do not allow me to analyze the possibility that learning increases future earnings, not current ones.[42]

7 Conclusion

Cities are characterized by workers' high skill levels, and workers can improve their skills through learning. This paper presents new stylized facts about learning behavior in cities. I directly examine off-the-job learning behavior using a Japanese survey that contains very detailed information on the frequency, purpose, and subject of learning. Although learning is generally considered to be more facilitated in cities, most empirical studies on the urban wage premium have looked at wage dynamics but have not directly analyzed learning behavior (e.g., Glaeser and Maré 2001; Baum-Snow and Pavan 2012; De la Roca and Puga 2017). This paper directly investigates learning behavior by exploiting a distinctive survey on off-the-job learning. This allows me to analyze the frequency, purpose, and subject of learning, which are hard to infer from wage data.

I document three stylized facts. First, people in denser cities engage in off-the-job learning more frequently: a 1% increase in population density is associated with a 0.1% increase in learning frequency. This suggests that the faster wage growth in cities might be partially attributed to greater efforts by urban residents. This provides a new microfoundation for urban agglomeration economies: productivity in denser cities is higher because people voluntarily improve their skills. Moreover, off-the-job learning might also provide a microfoundation for the comparative advantage of cities in skill-intensive industries, beyond the sorting of skilled workers into cities (Davis and Dingel 2020).

Second, people in denser cities are more likely to learn to obtain a new job or for self-improvement. This is consistent with the view that there are more job opportunities in denser cities (Di Addario 2011; Papageorgiou 2022) and that people have stronger incentives to fulfill the skill requirements of jobs (c.f., Autor et al. 2003; Ikenaga and Kambayashi 2016), which are more demanding in larger cities (Atalay et al. 2024). Learning for self-improvement implies greater engagement in cultural activities in denser cities and is consistent with the agglomeration

[41] An alternative method leads to the similar number. I have $\gamma_1 = 0.038$ after controlling for the learning frequency. Without controlling for it, the estimate is 0.039 (see Column 1 of Table E.1). Thus, about (0.039-0.038)/0.039 = 2.5% of the urban earnings premium is mediated through learning frequency.

[42] In particular, this paper focuses on active off-the-job learning that resembles an investment that may bring benefits in the future (e.g., finding a new job or getting promoted) at the cost of current time and effort costs. Moreover, such learning may be particularly active when a negative earnings shock hits, like the dip in job training programs (Ashenfelter 1978). These factors mask the positive effect of learning on income by creating a negative correlation between current learning frequency and income. Unfortunately, my cross-sectional data do not allow me to further investigate these possibilities.

of cultural activities or creative industries (e.g., Mitchell 2019; Tao et al. 2019; Borowiecki and Dahl 2021). This might also be related to studies on endogenous amenities (e.g., Shapiro 2006; Diamond 2016) because cultural activities might improve neighborhood quality through cultural amenities or the cultural sophistication of residents.

Third, people in denser cities are generally more likely to learn skills related to skill-intensive sectors, such as foreign languages and computer literacy. Therefore, off-the-job learning in cities might indeed reinforce their comparative advantage in skill-intensive sectors. This is also consistent with empirical findings that jobs in larger cities require greater proficiency in computer skills and social skills (Atalay et al. 2024). Interestingly, elderly care skills, which are more physically demanding, tend to be learned in less dense cities. This is consistent with the high demand for elderly care in rural Japanese labor markets, where population aging is particularly severe. Overall, people appear to respond to local demand conditions for skills, whereby cities face a higher demand for high-skilled white-collar jobs, while rural areas may face a higher demand for jobs that require physical skills, such as elderly care.

I believe that these new stylized facts will guide further theoretical and empirical investigations into learning and cities, providing valuable insights into urban agglomeration economies and the comparative advantages of cities in skill-intensive sectors. That said, certain limitations of my findings should be considered, which future work may hopefully overcome. First, further research is needed to move beyond descriptive evidence toward causal statements. Although my analyses in Section 4 suggest that my estimates might approximate causality, obtaining richer panel data or leveraging a natural experiment could further bolster causal inference. Second, I lack data on on-the-job learning. People may improve their skills in the workplace through communication with others (e.g., Charlot and Duranton 2004, 2006; Sandvik et al. 2020). While my results might also be suggestive of on-the-job learning, its direct investigation remains an important task. Finally, it would be valuable to better disentangle the driving forces behind learning behaviors. Although my discussion suggests that labor market considerations in denser cities can explain many of my key results, more detailed data on the motivation behind each learning behavior could further illuminate the underlying mechanisms.

References

Ahlfeldt, Gabriel M and Elisabetta Pietrostefani, "The economic effects of density: A synthesis," *Journal of Urban Economics*, 2019, *111*, 93–107.

Ashenfelter, Orley, "Estimating the effect of training programs on earnings," *The Review of Economics and Statistics*, 1978, pp. 47–57.

Atalay, Enghin, Sebastian Sotelo, and Daniel Tannenbaum, "The geography of job tasks," *Journal of Labor Economics*, 2024, *42*, 979–1008.

Autor, David H, Frank Levy, and Richard J Murnane, "The skill content of recent technological change: An empirical exploration," *The Quarterly Journal of Economics*, 2003, *118* (4), 1279–1333.

Azar, José, Ioana Marinescu, Marshall Steinbaum, and Bledi Taska, "Concentration in US labor markets: Evidence from online vacancy data," *Labour Economics*, 2020, *66*, 101886.

Bacolod, Marigee, Bernardo S Blum, and William C Strange, "Skills in the city," *Journal of Urban Economics*, 2009, *65* (2), 136–153.

_ **, Bernardo S. Blum, Marcos A. Rangel, and William Strange**, "Learners in cities: Agglomeration and the spatial division of cognition," *Regional Science and Urban Economics*, 2023, *98*, 103838.

Baum-Snow, Nathaniel and Ronni Pavan, "Understanding the city size wage gap," *The Review of Economic Studies*, 2012, *79* (1), 88–127.

Behrens, Kristian, Gilles Duranton, and Frédéric Robert-Nicoud, "Productive cities: Sorting, selection, and agglomeration," *Journal of Political Economy*, 2014, *122* (3), 507–553.

Bellemare, Marc F and Casey J Wichman, "Elasticities and the inverse hyperbolic sine transformation," *Oxford Bulletin of Economics and Statistics*, 2020, *82* (1), 50–61.

Borowiecki, Karol Jan and Christian Møller Dahl, "What makes an artist? The evolution and clustering of creative activity in the US since 1850," *Regional Science and Urban Economics*, 2021, *86*, 103614.

Boualam, Brahim, "Does culture affect local productivity and urban amenities?," *Regional Science and Urban Economics*, 2014, *46*, 12–17.

Charlot, Sylvie and Gilles Duranton, "Communication externalities in cities," *Journal of Urban Economics*, 2004, *56* (3), 581–613.

_ **and** _ , "Cities and workplace communication: some quantitative French evidence," *Urban Studies*, 2006, *43* (8), 1365–1394.

Ciccone, Antonio and Robert E Hall, "Productivity and the Density of Economic Activity," *American Economic Review*, 1996, *86* (1), 54–70.

Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon, "The identification of agglomeration economies," *Journal of Economic Geography*, 2011, *11* (2), 253–266.

Dauth, Wolfgang, Sebastian Findeisen, Enrico Moretti, and Jens Suedekum, "Matching in Cities," *Journal of the European Economic Association*, 2022, *20* (4), 1478–1521.

Davis, Donald R and Jonathan I Dingel, "A spatial knowledge economy," *American Economic Review*, 2019, *109* (1), 153–70.

_ **and** _ , "The comparative advantage of cities," *Journal of International Economics*, 2020, *123*, 103291.

De la Roca, Jorge and Diego Puga, "Learning by working in big cities," *The Review of Economic Studies*, 2017, *84* (1), 106–142.

— **, Gianmarco IP Ottaviano, and Diego Puga**, "City of dreams," *Journal of the European Economic Association*, 2023, *21*, 690–726.

Di Addario, Sabrina, "Job search in thick markets," *Journal of Urban Economics*, 2011, *69* (3), 303–318.

Diamond, Rebecca, "The determinants and welfare implications of US workers' diverging location choices by skill: 1980-2000," *American Economic Review*, 2016, *106* (3), 479–524.

Duranton, Gilles and Diego Puga, "Micro-foundations of urban agglomeration economies," in "Handbook of Regional and Urban Economics," Vol. 4, Elsevier, 2004, pp. 2063–2117.

Eckert, Fabian, Mads Hejlesen, and Conor Walsh, "The return to big-city experience: Evidence from refugees in Denmark," *Journal of Urban Economics*, 2022, *130*, 103454.

Gaulke, Amanda P, "Post-schooling off-the-job training and its benefits," *Labour Economics*, 2021, *70*, 102007.

Glaeser, Edward L, "Learning in cities," *Journal of Urban Economics*, 1999, *46* (2), 254–277.

— **and David C Maré**, "Cities and skills," *Journal of Labor Economics*, 2001, *19* (2), 316–342.

Hanaoka, Chie, "Kaigo roudou-ryoku busoku ha naze shouzite irunoka (Direct Care Worker Shortages Using Japanese Data : A Review of Previous Studies)," *Nihon Roudou Kenkyu Zasshi*, 2015, *57* (5), 16–25. (In Japanese).

Ikenaga, Toshie and Ryo Kambayashi, "Task Polarization in the Japanese Labor Market: Evidence of a Long-Term Trend," *Industrial Relations*, 2016, *55* (2), 267–293.

Kanemoto, Yoshitsugu and Kazuyuki Tokuoka, "Nihon no toshi-ken settei kijun (Proposal for the standards of metropolitan areas of Japan)," *Journal of Applied Regional Science*, 2002, *7*, 1–15. (In Japanese. An English abstract is available).

Kawaguchi, Daiji, "Applying the Mincer Wage Equation to Japanese Data (*Mincer gata chingin kansu no nihon no roudou shijyo heno ouyou* in Japanese)," *RIETI Discussion Paper Series*, 2011, *11-J-026*, 1–26.

Kondo, Keisuke, "Shicho-son gappei wo kouryo shita shikuchoson panel deta no sakusei (Municipality-level Panel Data and Municipal Mergers in Japan)," 2022. RIETI Technical Paper Series 19-T-001. (In Japanese. An English abstract is available).

Krupka, Douglas J, "Location-specific human capital, location choice and amenity demand," *Journal of Regional Science*, 2009, *49* (5), 833–854.

Kuroda, Sachiko, "Do Japanese work shorter hours than before? Measuring trends in market work and leisure using 1976–2006 Japanese time-use survey," *Journal of the Japanese and International Economies*, 2010, *24* (4), 481–502.

Lee, Jungmin, Daiji Kawaguchi, and Daniel S Hamermesh, "Aggregate impacts of a gift of time," *American Economic Review*, 2012, *102* (3), 612–16.

Lynch, Lisa M, "Private-sector training and the earnings of young workers," *American Economic Review*, 1992, *82* (1), 299–312.

Marshall, Alfred, *Principles of Economics*, MacMillan, 1890.

Mitchell, Sara, "London calling? Agglomeration economies in literature since 1700," *Journal of Urban Economics*, 2019, *112*, 16–32.

Miyauchi, Yuhei, "Matching and agglomeration: Theory and evidence from Japanese firm-to-firm trade," *Econometrica*, 2024, *92*, 1868–1905.

Murphy, Daniel, "Home production, expenditure, and economic geography," *Regional Science and Urban Economics*, 2018, *70*, 112–126.

Oster, Emily, "Unobservable selection and coefficient stability: Theory and evidence," *Journal of Business & Economic Statistics*, 2019, *37* (2), 187–204.

Papageorgiou, Theodore, "Learning your comparative advantages," *Review of Economic Studies*, 2014, *81* (3), 1263–1295.

_ , "Occupational Matching and Cities," *American Economic Journal: Macroeconomics*, 2022, *14* (3), 82–132.

Rosenthal, Stuart S and William C Strange, "Evidence on the nature and sources of agglomeration economies," in "Handbook of Regional and Urban Economics," Vol. 4, Elsevier, 2004, pp. 2119–2171.

Sandvik, Jason J, Richard E Saouma, Nathan T Seegert, and Christopher T Stanton, "Workplace knowledge flows," *The Quarterly Journal of Economics*, 2020, *135* (3), 1635–1680.

Santos Silva, J.M.C. and Silvana Tenreyro, "The log of gravity," *The Review of Economics and Statistics*, 2006, *88* (4), 641–658.

Shapiro, Jesse M, "Smart cities: quality of life, productivity, and the growth effects of human capital," *The Review of Economics and Statistics*, 2006, *88* (2), 324–335.

Su, Yichen, "Measuring the value of urban consumption amenities: A time-use approach," *Journal of Urban Economics*, 2022, *132*, 103495.

Tabuchi, Takatoshi and Atsushi Yoshida, "Separating urban agglomeration economies in consumption and production," *Journal of Urban Economics*, 2000, *48* (1), 70–84.

Tao, Jin, Chun-Yu Ho, Shougui Luo, and Yue Sheng, "Agglomeration economies in creative industries," *Regional Science and Urban Economics*, 2019, *77*, 141–154.

Appendix to "Off-the-Job Learning in Cities"
(Not for Publication)

A Omitted figures and tables

	mean	sd	count
Number of days learning	39.34	113.63	79879
Dummy of learning something	0.38	0.49	79879
Dummy of learning for a new job	0.084	0.278	25170
Dummy of learning for a current job	0.510	0.500	25170
Dummy of learning for cultivation	0.560	0.496	24729
Dummy of learning for other objectives	0.260	0.439	25170
Dummy of learning English	0.110	0.313	79879
Dummy of learning other foreign language	0.035	0.183	79879
Dummy of learning computers etc	0.132	0.338	79879
Dummy of learning commerce and business	0.104	0.306	79879
Dummy of learning elderly care	0.039	0.193	79879
Dummy of learning housekeeping	0.106	0.307	79879
Dummy of learning humanities and sciences	0.074	0.262	79879
Dummy of learning arts and culture	0.100	0.300	79879
Dummy of learning others	0.079	0.270	79879
Dummy of learning by classes (public)	0.046	0.209	24553
Dummy of learning by classes (private)	0.146	0.353	25170
Dummy of learning by classes (universities etc)	0.019	0.137	25170
Dummy of learning by lecture meetings	0.098	0.297	25170
Dummy of learning by home-study courses	0.074	0.262	25170
Dummy of learning by TV and radio	0.193	0.394	25170
Dummy of learning by workplace outside working hours	0.235	0.424	25170
Dummy of learning by vocational schools etc	0.023	0.149	25170
Dummy of learning by training by employment service center	0.011	0.105	25170
Dummy of learning by others	0.588	0.489	25170

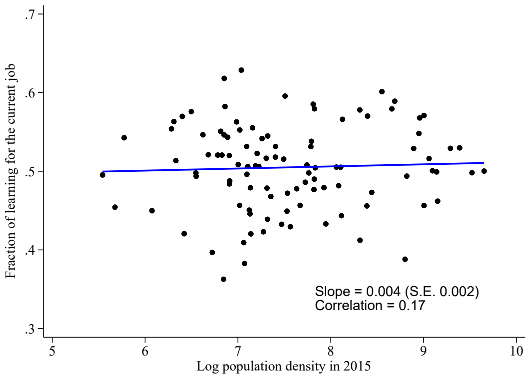
Table A.1: Summary statistics on learning variables

Note: All variables are weighted by the sampling weights. The number of observations for a purpose or a method of learning is smaller because they are defined only for those learning something. The dummy of learning for cultivation has slightly fewer observations than the other dummy variables on purposes because some observations are coded as “unknown" in the original data. Similarly, the dummy of learning by classes (public) has slightly fewer observations than the other dummy variables on methods because some observations are coded as “unknown" in the original data.

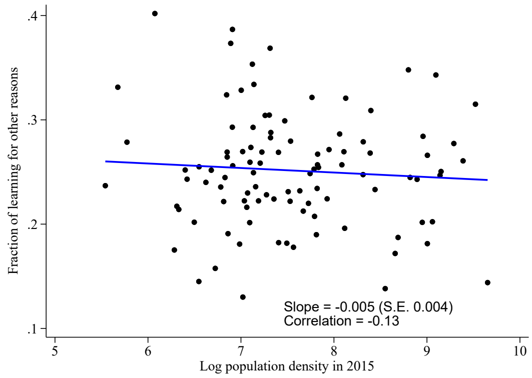
	mean	sd	count
Learners (those who reported some learning behavior)			
Age	42.11	9.85	23744
Dummy for female	0.506	0.500	23744
Dummy for being married	0.670	0.472	23744
Dummy for college or more	0.471	0.499	23566
Individual annual income (10,000 Japanese yen)	428.62	323.12	20525
Non-learners (those who did not report any learning behavior)			
Age	42.80	9.43	56135
Dummy for female	0.491	0.500	56135
Dummy for being married	0.671	0.470	56135
Dummy for college or more	0.259	0.438	53406
Individual annual income (10,000 Japanese yen)	358.24	263.94	47454
Learners & Non-learners			
Age	42.57	9.58	79879
Dummy for female	0.496	0.500	79879
Dummy for being married	0.669	0.471	79879
Dummy for college or more	0.330	0.470	78972
Individual annual income (10,000 Japanese yen)	382.22	283.63	67979
Distribution of municipal population density			
ln 2015 municipal population density	6.630	1.209	1350 (the number of municipalities)

Table A.2: Summary statistics on individual characteristics

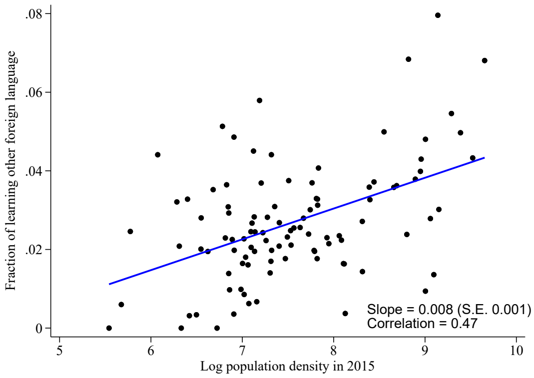
Note: Learners are defined as individuals who engaged in some learning behavior in a year (i.e., the number of days spent on learning is larger than zero). Similarly, non-learners are defined as individuals who did not engage in any learning behavior in a year (i.e., the number of days spent on learning equals zero) All individual characteristics are weighted by the sampling weights. In the data, marital status take three values (unmarried, married, widowed) and education attainment takes nine values (elementary school, junior high school, high school, vocational school (1–2 years), vocational school (2–4 years), vocational school (more than 4 years), associate degree, bachelors degree, and master or doctoral degree. I summarize these two variables into binary dummies to concisely present the summary statistics. The number of observations varies across characteristics due to missing values. The summary statistics of the log population density is calculated, without weights, at the municipal level because this is the level of variation.



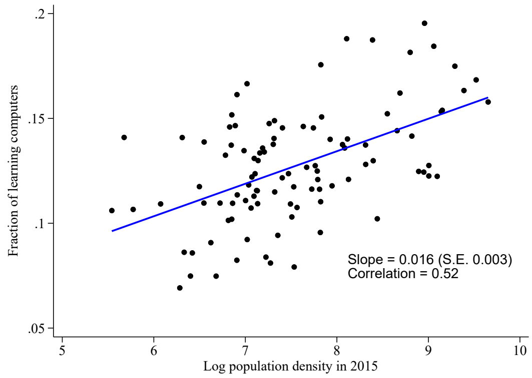
(a) Fraction of learning for the current job



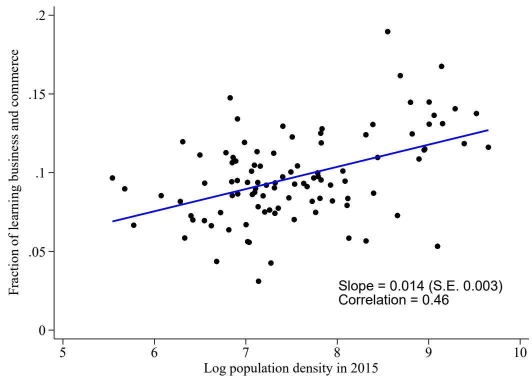
(b) Fraction of learning for other reasons



(c) Fraction of learning other foreign language



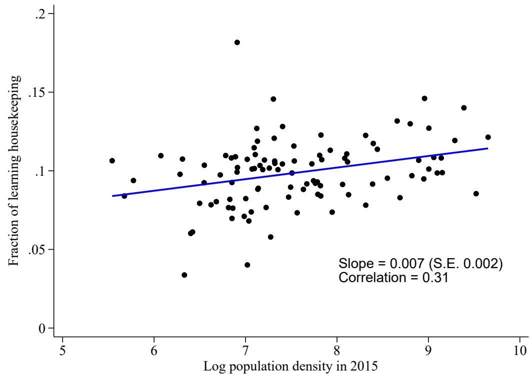
(d) Fraction of learning computers etc



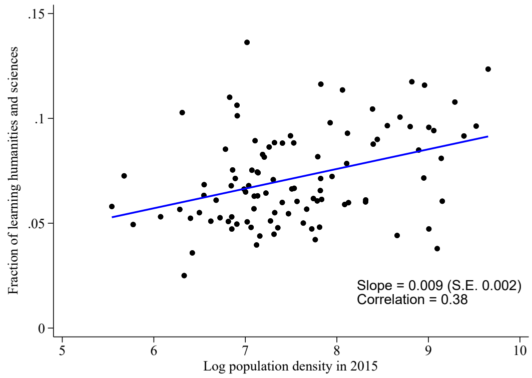
(e) Fraction of learning business and commerce

Figure A.1: Learning outcomes and population density

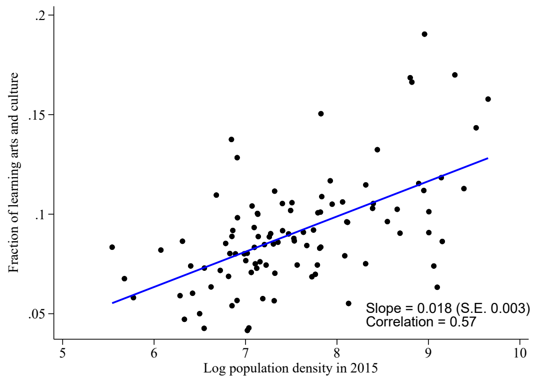
Note: The figures plot the municipal average of learning outcomes against its log population density in 2015 (black dots). The blue line is the regression line. The fractions for a purpose or a method of learning are calculated among those who learned something. To credibly calculate municipal averages and to comply with the data privacy concern by the data provider, I focus on municipalities with more than 150 individual observations (101 municipalities) in these figures. Using alternative threshold numbers of minimum observations produces similar figures.



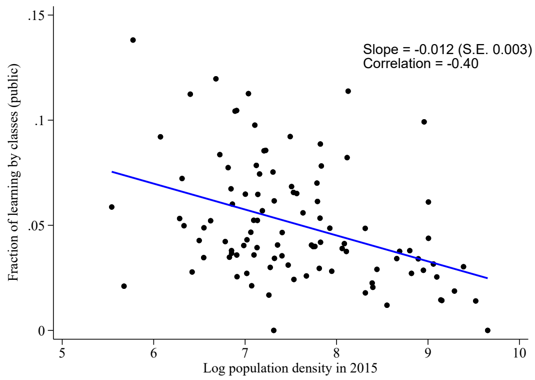
(f) Fraction of learning housekeeping



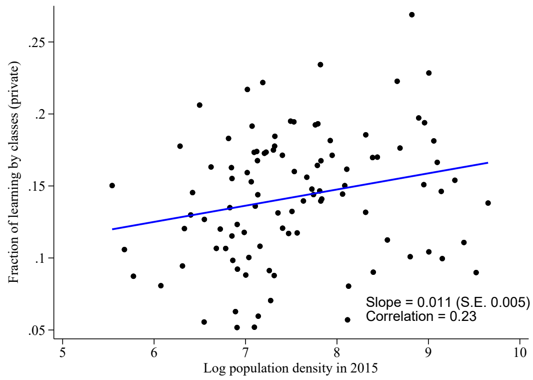
(g) Fraction of learning humanities and social sciences



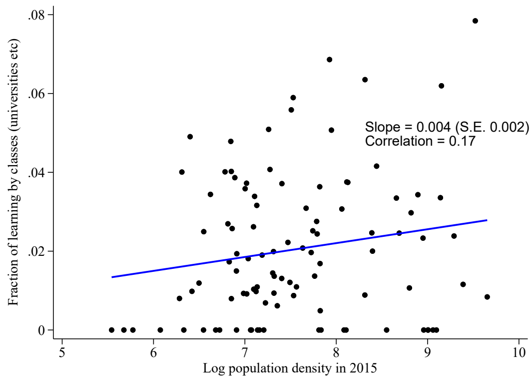
(h) Fraction of learning arts and sciences



(i) Fraction of learning by classes (public)



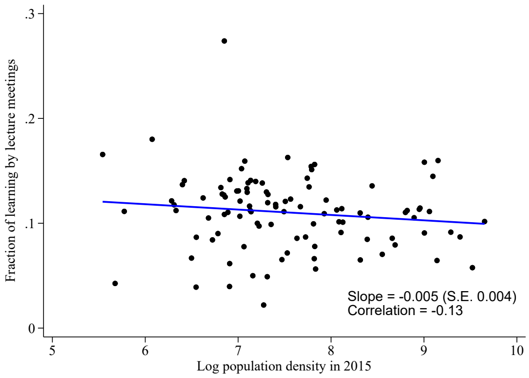
(j) Fraction of learning by classes (private)



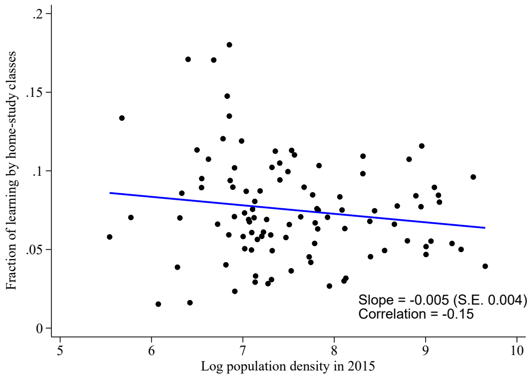
(k) Fraction of learning by classes (universities etc)

Figure A.1: Learning outcomes and population density (cont.)

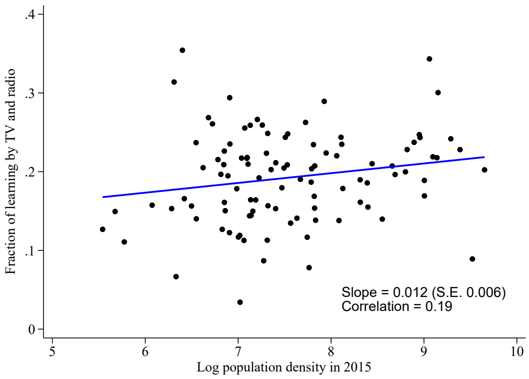
Note: The figures plot the municipal average of learning outcomes against its log population density in 2015 (black dots). The blue line is the regression line. The fractions for a purpose or a method of learning are calculated among those who learned something. To credibly calculate municipal averages and to comply with the data privacy concern by the data provider, I focus on municipalities with more than 150 individual observations (101 municipalities) in these figures. Using alternative threshold numbers of minimum observations produces similar figures.



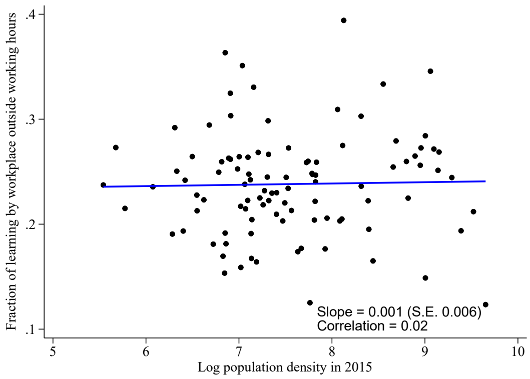
(l) Fraction of learning by lecture meetings



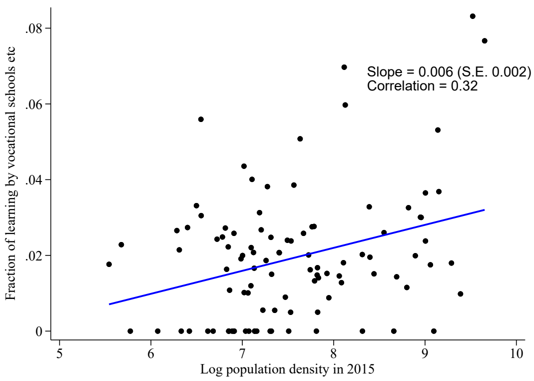
(m) Fraction of learning by home-study classes



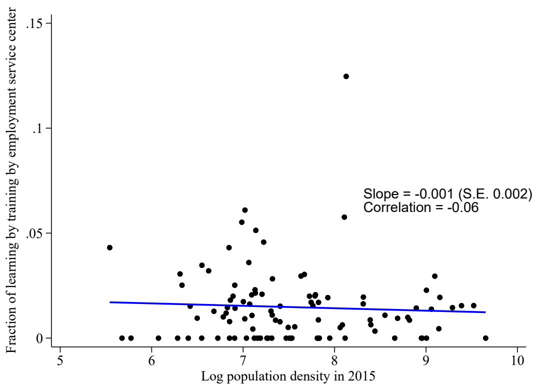
(n) Fraction of learning by TV and radio



(o) Fraction of learning by workplace outside working hours



(p) Fraction of learning by vocational schools etc



(q) Fraction of learning by training at employment service

Figure A.1: Learning outcomes and population density (cont.)

Note: The figures plot the municipal average of learning outcomes against its log population density in 2015 (black dots). The blue line is the regression line. The fractions for a purpose or a method of learning are calculated among those who learned something. To credibly calculate municipal averages and to comply with the data privacy concern by the data provider, I focus on municipalities with more than 150 individual observations (101 municipalities) in these figures. Using alternative threshold numbers of minimum observations produces similar figures.

	Number of days learning (1)	Log number of days learning (2)	Engaged in learning (3)	
ln population density	0.089^{a}	0.046^{b}	0.020^{a}	
	(0.019)	(0.019)	(0.003)	
Estimation method	PPML	OLS	OLS	
Oster's bound	NA	0.039	0.011	
N	67650	20446	67650	
R^2	NA	0.018	0.089	
	Learning for new employment (4)	Learning for current job (5)	Learning for cultivation (6)	Learning for other objectives (7)
ln population density	0.008^{b}	-0.004	0.023^{a}	0.003
	(0.004)	(0.004)	(0.005)	(0.005)
Estimation method	OLS	OLS	OLS	OLS
Oster's bound	0.008	-0.008	0.020	0.006
N	21551	21551	21210	21551
R^2	0.012	0.077	0.033	0.027

Cluster-robust standard errors in parentheses

^{c} $p < 0.1$, ^{b} $p < 0.05$, ^{a} $p < 0.01$

Table A.3: Learning frequency and purposes (controlling for occupation)

Note: In Column 1, I estimate by the Pseudo Poisson Maximum Likelihood (PPML, Santos Silva and Tenreyro 2006) the constant-elasticity regression model (2) relating number of days spent for learning to the municipal population density. Column 2 takes as the outcome variable the log number of days spent for learning in the linear regression model (1). The remaining columns use the linear probability model as the outcome variable is a dummy. Column 6 has slightly fewer observations than columns 4, 5, and 7 because some observations are coded as "unknown." In all specifications, I include as control variables sex, age and its square, marital status, educational attainment, and occupation. I calculate Oster's (2019) beta for each regression. Based on a recommendation of Oster (2019), I set the maximal R^2 to 1.3 times the R^2 of each regression. I also assume that observed and unobserved factors are equally related to population density (i.e., the coefficient of proportionality $\delta = 1$ in Oster's notation). Due to its availability, I do not report it when using the PPML. I cluster standard errors at the municipal level.

	English	Other foreign language	Computers etc	Commerce and Business	Elderly care
	(1)	(2)	(3)	(4)	(5)
ln population density	0.019[a]	0.006[a]	0.008[a]	0.009[a]	-0.004[a]
	(0.002)	(0.001)	(0.002)	(0.002)	(0.001)
Estimation method	OLS	OLS	OLS	OLS	OLS
Oster's bound	0.014	0.005	0.005	0.005	-0.004
N	67650	67650	67650	67650	67650
R^2	0.087	0.016	0.036	0.056	0.048
	Housekeeping	Humanities and sciences	Arts and culture	Other	
	(6)	(7)	(8)	(9)	
ln population density	0.008[a]	0.003[b]	0.012[a]	0.000	
	(0.002)	(0.002)	(0.002)	(0.002)	
Estimation method	OLS	OLS	OLS	OLS	
Oster's bound	0.007	0.000	0.009	-0.001	
N	67650	67650	67650	67650	
R^2	0.039	0.059	0.046	0.013	

Cluster-robust standard errors in parentheses

[c] $p < 0.1$, [b] $p < 0.05$, [a] $p < 0.01$

Table A.4: Learning by subject (controlling for occupation)

Note: I use the linear probability model that regresses the dummy for learning a particular subject on the municipal population density. In all specifications, I include as control variables sex, age and its square, marital status, educational attainment, and occupation. I calculate Oster's (2019) beta for each regression. Based on a recommendation of Oster (2019), I set the maximal R^2 to 1.3 times the R^2 of each regression. I also assume that observed and unobserved factors are equally related to population density (i.e., the coefficient of proportionality $\delta = 1$ in Oster's notation). I cluster standard errors at the municipal level.

	Number of days learning	Log number of days learning	Engaged in learning	
	(1)	(2)	(3)	
ln population density	0.086^{a}	0.044^{a}	0.016^{a}	
	(0.018)	(0.017)	(0.003)	
Estimation method	PPML	OLS	OLS	
Oster's bound	NA	0.032	0.005	
N	78727	23513	78727	
R^2	NA	0.015	0.070	
	Learning for new employment	Learning for current job	Learning for cultivation	Learning for other objectives
	(4)	(5)	(6)	(7)
ln population density	0.007	-0.009^{b}	0.024^{a}	0.003
	(0.004)	(0.005)	(0.005)	(0.005)
Estimation method	OLS	OLS	OLS	OLS
Oster's bound	0.006	-0.015	0.019	0.006
N	24905	24905	24475	24905
R^2	0.017	0.070	0.032	0.025

Cluster-robust standard errors in parentheses

^{c} $p < 0.1$, ^{b} $p < 0.05$, ^{a} $p < 0.01$

Table A.5: Learning frequency and purposes (controlling for industry concentration)

Note: In Column 1, I estimate by the Pseudo Poisson Maximum Likelihood (PPML, Santos Silva and Tenreyro 2006) the constant-elasticity regression model (2) relating number of days spent for learning to the municipal population density. Column 2 takes as the outcome variable the log number of days spent for learning in the linear regression model (1). The remaining columns use the linear probability model as the outcome variable is a dummy. Column 6 has slightly fewer observations than columns 4, 5, and 7 because some observations are coded as "unknown." In all specifications, I include as control variables sex, age and its square, marital status, educational attainment, and the Herfindahl-Hirschman Index of industry concentration. I calculate Oster's (2019) beta for each regression. Based on a recommendation of Oster (2019), I set the maximal R^2 to 1.3 times the R^2 of each regression. I also assume that observed and unobserved factors are equally related to population density (i.e., the coefficient of proportionality $\delta = 1$ in Oster's notation). Due to its availability, I do not report it when using the PPML. I cluster standard errors at the municipal level.

	English	Other foreign language	Computers etc	Commerce and Business	Elderly care
	(1)	(2)	(3)	(4)	(5)
ln population density	0.019[a]	0.007[a]	0.007[a]	0.008[a]	-0.004[a]
	(0.002)	(0.001)	(0.002)	(0.002)	(0.001)
Estimation method	OLS	OLS	OLS	OLS	OLS
Oster's bound	0.013	0.005	0.003	0.003	-0.005
N	78727	78727	78727	78727	78727
R^2	0.080	0.016	0.027	0.039	0.013
	Housekeeping	Humanities and sciences	Arts and culture	Other	
	(6)	(7)	(8)	(9)	
ln population density	0.007[a]	0.002	0.012[a]	-0.000	
	(0.002)	(0.002)	(0.002)	(0.002)	
Estimation method	OLS	OLS	OLS	OLS	
Oster's bound	0.006	-0.002	0.008	-0.002	
N	78727	78727	78727	78727	
R^2	0.036	0.050	0.038	0.007	

Cluster-robust standard errors in parentheses
^{c} $p < 0.1$, ^{b} $p < 0.05$, ^{a} $p < 0.01$

Table A.6: Learning by subject (controlling for industry concentration)

Note: I use the linear probability model that regresses the dummy for learning a particular subject on the municipal population density. In all specifications, I include as control variables sex, age and its square, marital status, educational attainment, and the Herfindahl-Hirschman Index of industry concentration. I calculate Oster's (2019) beta for each regression. Based on a recommendation of Oster (2019), I set the maximal R^2 to 1.3 times the R^2 of each regression. I also assume that observed and unobserved factors are equally related to population density (i.e., the coefficient of proportionality $\delta = 1$ in Oster's notation). I cluster standard errors at the municipal level.

	Number of days learning	Log number of days learning	Engaged in learning	
	(1)	(2)	(3)	
ln population density	0.077^a	0.071^a	0.025^a	
	(0.024)	(0.026)	(0.005)	
Estimation method	PPML	OLS	OLS	
Oster's bound	NA	0.071	0.024	
N	20472	9169	20472	
R^2	NA	0.016	0.024	
	Learning for new employment	Learning for current job	Learning for cultivation	Learning for other objectives
	(4)	(5)	(6)	(7)
ln population density	0.001	-0.002	0.015^c	0.011
	(0.005)	(0.007)	(0.008)	(0.008)
Estimation method	OLS	OLS	OLS	OLS
Oster's bound	0.001	-0.002	0.015	0.011
N	9548	9548	9457	9548
R^2	0.019	0.048	0.030	0.011

Cluster-robust standard errors in parentheses

^{c} $p < 0.1$, ^{b} $p < 0.05$, ^{a} $p < 0.01$

Table A.7: Learning frequency and purposes (college-graduates sample)

Note: In Column 1, I estimate by the Pseudo Poisson Maximum Likelihood (PPML, Santos Silva and Tenreyro 2006) the constant-elasticity regression model (2) relating number of days spent for learning to the municipal population density. Column 2 takes as the outcome variable the log number of days spent for learning in the linear regression model (1). The remaining columns use the linear probability model as the outcome variable is a dummy. Column 6 has slightly fewer observations than columns 4, 5, and 7 because some observations are coded as "unknown." In all specifications, I include as control variables sex, age and its square, marital status, and educational attainment. I calculate Oster's (2019) beta for each regression. Based on a recommendation of Oster (2019), I set the maximal R^2 to 1.3 times the R^2 of each regression. I also assume that observed and unobserved factors are equally related to population density (i.e., the coefficient of proportionality $\delta = 1$ in Oster's notation). Due to its availability, I do not report it when using the PPML. I cluster standard errors at the municipal level.

	English	Other foreign language	Computers etc	Commerce and Business	Elderly care
	(1)	(2)	(3)	(4)	(5)
ln population density	0.027[a]	0.009[a]	0.008[b]	0.013[a]	-0.002
	(0.004)	(0.003)	(0.004)	(0.005)	(0.002)
Estimation method	OLS	OLS	OLS	OLS	OLS
Oster's bound	0.026	0.009	0.008	0.013	-0.002
N	20472	20472	20472	20472	20472
R^2	0.043	0.013	0.016	0.021	0.006
	Housekeeping	Humanities and sciences	Arts and culture	Other	
	(6)	(7)	(8)	(9)	
ln population density	0.008[c]	0.000	0.013[a]	-0.003	
	(0.004)	(0.003)	(0.004)	(0.003)	
Estimation method	OLS	OLS	OLS	OLS	
Oster's bound	0.007	-0.000	0.012	0.003	
N	20472	20472	20472	20472	
R^2	0.040	0.029	0.031	0.003	

Cluster-robust standard errors in parentheses
[c] $p < 0.1$, [b] $p < 0.05$, [a] $p < 0.01$

Table A.8: Learning by subject (college-graduates sample)

Note: I use the linear probability model that regresses the dummy for learning a particular subject on the municipal population density. In all specifications, I include as control variables sex, age and its square, marital status, and educational attainment. I calculate Oster's (2019) beta for each regression. Based on a recommendation of Oster (2019), I set the maximal R^2 to 1.3 times the R^2 of each regression. I also assume that observed and unobserved factors are equally related to population density (i.e., the coefficient of proportionality $\delta = 1$ in Oster's notation). I cluster standard errors at the municipal level.

	Number of days learning (1)	Log number of days learning (2)	Engaged in learning (3)	
ln population density	0.124[a]	0.043[b]	0.019[a]	
	(0.021)	(0.018)	(0.003)	
Estimation method	PPML	OLS	OLS	
Oster's bound	NA	0.043	0.017	
N	58255	14344	58255	
R^2	NA	0.007	0.028	
	Learning for new employment (4)	Learning for current job (5)	Learning for cultivation (6)	Learning for other objectives (7)
ln population density	0.013[b]	-0.014[b]	0.033[a]	-0.001
	(0.006)	(0.006)	(0.005)	(0.006)
Estimation method	OLS	OLS	OLS	OLS
Oster's bound	0.013	-0.014	0.031	-0.001
N	15347	15347	15018	15347
R^2	0.017	0.053	0.025	0.015

Cluster-robust standard errors in parentheses

^{c} $p < 0.1$, ^{b} $p < 0.05$, ^{a} $p < 0.01$

Table A.9: Learning frequency and purposes (non-college-graduates sample)

Note: In Column 1, I estimate by the Pseudo Poisson Maximum Likelihood (PPML, Santos Silva and Tenreyro 2006) the constant-elasticity regression model (2) relating number of days spent for learning to the municipal population density. Column 2 takes as the outcome variable the log number of days spent for learning in the linear regression model (1). The remaining columns use the linear probability model as the outcome variable is a dummy. Column 6 has slightly fewer observations than columns 4, 5, and 7 because some observations are coded as "unknown." In all specifications, I include as control variables sex, age and its square, marital status, and educational attainment. I calculate Oster's (2019) beta for each regression. Based on a recommendation of Oster (2019), I set the maximal R^2 to 1.3 times the R^2 of each regression. I also assume that observed and unobserved factors are equally related to population density (i.e., the coefficient of proportionality $\delta = 1$ in Oster's notation). Due to its availability, I do not report it when using the PPML. I cluster standard errors at the municipal level.

	English	Other foreign language	Computers etc	Commerce and Business	Elderly care
	(1)	(2)	(3)	(4)	(5)
ln population density	0.017[a]	0.006[a]	0.010[a]	0.009[a]	-0.004[a]
	(0.002)	(0.001)	(0.002)	(0.002)	(0.001)
Estimation method	OLS	OLS	OLS	OLS	OLS
Oster's bound	0.016	0.006	0.001	0.009	-0.005
N	58255	58255	58255	58255	58255
R^2	0.020	0.006	0.013	0.012	0.016
	Housekeeping	Humanities and sciences	Arts and culture	Other	
	(6)	(7)	(8)	(9)	
ln population density	0.007[a]	0.006[a]	0.014[a]	0.002	
	(0.002)	(0.002)	(0.002)	(0.002)	
Estimation method	OLS	OLS	OLS	OLS	
Oster's bound	0.006	0.006	0.013	0.001	
N	58255	58255	58255	58255	
R^2	0.034	0.007	0.002	0.007	

Cluster-robust standard errors in parentheses
^{c} $p < 0.1$, ^{b} $p < 0.05$, ^{a} $p < 0.01$

Table A.10: Learning by subject (non-college-graduates sample)

Note: I use the linear probability model that regresses the dummy for learning a particular subject on the municipal population density. In all specifications, I include as control variables sex, age and its square, marital status, and educational attainment. I calculate Oster's (2019) beta for each regression. Based on a recommendation of Oster (2019), I set the maximal R^2 to 1.3 times the R^2 of each regression. I also assume that observed and unobserved factors are equally related to population density (i.e., the coefficient of proportionality $\delta = 1$ in Oster's notation). I cluster standard errors at the municipal level.

	Number of days learning (1)	Log number of days learning (2)	Engaged in learning (3)	
ln population density	0.101[a]	0.040[b]	0.024[a]	
	(0.022)	(0.020)	(0.003)	
Estimation method	PPML	OLS	OLS	
Oster's bound	NA	0.029	0.016	
N	56984	17220	56984	
R^2	NA	0.019	0.075	
	Learning for new employment (4)	Learning for current job (5)	Learning for cultivation (6)	Learning for other objectives (7)
ln population density	0.008[c]	-0.005	0.023[a]	0.007
	(0.004)	(0.005)	(0.006)	(0.005)
Estimation method	OLS	OLS	OLS	OLS
Oster's bound	0.009	-0.009	0.020	0.009
N	18102	18102	17842	18102
R^2	0.012	0.035	0.028	0.015

Cluster-robust standard errors in parentheses

[c] $p < 0.1$, [b] $p < 0.05$, [a] $p < 0.01$

Table A.11: Learning frequency and purposes (people mainly devoted to work)

Note: In Column 1, I estimate by the Pseudo Poisson Maximum Likelihood (PPML, Santos Silva and Tenreyro 2006) the constant-elasticity regression model (2) relating number of days spent for learning to the municipal population density. Column 2 takes as the outcome variable the log number of days spent for learning in the linear regression model (1). The remaining columns use the linear probability model as the outcome variable is a dummy. Column 6 has slightly fewer observations than columns 4, 5, and 7 because some observations are coded as "unknown." In all specifications, I include as control variables sex, age and its square, marital status, and educational attainment. I calculate Oster's (2019) beta for each regression. Based on a recommendation of Oster (2019), I set the maximal R^2 to 1.3 times the R^2 of each regression. I also assume that observed and unobserved factors are equally related to population density (i.e., the coefficient of proportionality $\delta = 1$ in Oster's notation). Due to its availability, I do not report it when using the PPML. I cluster standard errors at the municipal level.

	English	Other foreign language	Computers etc	Commerce and Business	Elderly care
	(1)	(2)	(3)	(4)	(5)
ln population density	0.020[a]	0.007[a]	0.010[a]	0.014[a]	-0.004
	(0.002)	(0.002)	(0.002)	(0.003)	(0.001)
Estimation method	OLS	OLS	OLS	OLS	OLS
Oster's bound	0.015	0.006	0.007	0.010	-0.004
N	56984	56984	56984	56984	56984
R^2	0.085	0.013	0.025	0.034	0.018
	Housekeeping	Humanities and sciences	Arts and culture	Other	
	(6)	(7)	(8)	(9)	
ln population density	0.007[a]	0.004[c]	0.012[a]	0.000	
	(0.002)	(0.002)	(0.002)	(0.002)	
Estimation method	OLS	OLS	OLS	OLS	
Oster's bound	0.006	0.000	0.010	0.000	
N	56984	56984	56984	56984	
R^2	0.029	0.052	0.036	0.007	

Cluster-robust standard errors in parentheses

^{c} $p < 0.1$, ^{b} $p < 0.05$, ^{a} $p < 0.01$

Table A.12: Learning by subject (people mainly devoted to work)

Note: I use the linear probability model that regresses the dummy for learning a particular subject on the municipal population density. In all specifications, I include as control variables sex, age and its square, marital status, and educational attainment. I calculate Oster's (2019) beta for each regression. Based on a recommendation of Oster (2019), I set the maximal R^2 to 1.3 times the R^2 of each regression. I also assume that observed and unobserved factors are equally related to population density (i.e., the coefficient of proportionality $\delta = 1$ in Oster's notation). I cluster standard errors at the municipal level.

	Number of days learning (1)	Log number of days learning (2)	Engaged in learning (3)	
ln population density	0.080^{a}	0.044^{c}	0.017^{a}	
	(0.029)	(0.026)	(0.005)	
Estimation method	PPML	OLS	OLS	
Oster's bound	NA	0.040	0.011	
N	21743	6293	21743	
R^2	NA	0.034	0.057	
	Learning for new employment (4)	Learning for current job (5)	Learning for cultivation (6)	Learning for other objectives (7)
ln population density	0.002	0.010	0.027^{a}	-0.015
	(0.008)	(0.010)	(0.007)	(0.010)
Estimation method	OLS	OLS	OLS	OLS
Oster's bound	0.001	0.009	0.023	-0.012
N	6803	6803	6633	6803
R^2	0.088	0.011	0.042	0.020

Cluster-robust standard errors in parentheses

^{c} $p < 0.1$, ^{b} $p < 0.05$, ^{a} $p < 0.01$

Table A.13: Learning frequency and purposes (people not mainly devoted to work)

Note: In Column 1, I estimate by the Pseudo Poisson Maximum Likelihood (PPML, Santos Silva and Tenreyro 2006) the constant-elasticity regression model (2) relating number of days spent for learning to the municipal population density. Column 2 takes as the outcome variable the log number of days spent for learning in the linear regression model (1). The remaining columns use the linear probability model as the outcome variable is a dummy. Column 6 has slightly fewer observations than columns 4, 5, and 7 because some observations are coded as "unknown." In all specifications, I include as control variables sex, age and its square, marital status, and educational attainment. I calculate Oster's (2019) beta for each regression. Based on a recommendation of Oster (2019), I set the maximal R^2 to 1.3 times the R^2 of each regression. I also assume that observed and unobserved factors are equally related to population density (i.e., the coefficient of proportionality $\delta = 1$ in Oster's notation). Due to its availability, I do not report it when using the PPML. I cluster standard errors at the municipal level.

	English	Other foreign language	Computers etc	Commerce and Business	Elderly care
	(1)	(2)	(3)	(4)	(5)
ln population density	0.020[a]	0.006[a]	0.009[b]	0.007[b]	-0.002
	(0.003)	(0.002)	(0.004)	(0.003)	(0.002)
Estimation method	OLS	OLS	OLS	OLS	OLS
Oster's bound	0.016	0.005	0.008	0.006	-0.002
N	21743	21743	21743	21743	21743
R^2	0.068	0.033	0.028	0.015	0.012
	Housekeeping	Humanities and sciences	Arts and culture	Other	
	(6)	(7)	(8)	(9)	
ln population density	0.005	0.005	0.015[a]	0.001	
	(0.005)	(0.003)	(0.004)	(0.003)	
Estimation method	OLS	OLS	OLS	OLS	
Oster's bound	0.004	0.003	0.012	-0.000	
N	21743	21743	21743	21743	
R^2	0.017	0.042	0.043	0.009	

Cluster-robust standard errors in parentheses

^{c} $p < 0.1$, ^{b} $p < 0.05$, ^{a} $p < 0.01$

Table A.14: Learning by subject (people not mainly devoted to work)

Note: I use the linear probability model that regresses the dummy for learning a particular subject on the municipal population density. In all specifications, I include as control variables sex, age and its square, marital status, and educational attainment. I calculate Oster's (2019) beta for each regression. Based on a recommendation of Oster (2019), I set the maximal R^2 to 1.3 times the R^2 of each regression. I also assume that observed and unobserved factors are equally related to population density (i.e., the coefficient of proportionality $\delta = 1$ in Oster's notation). I cluster standard errors at the municipal level.

	Number of days learning (1)	Log number of days learning (2)	Engaged in learning (3)	
ln population density	0.143[a]	0.098[b]	0.024[a]	
	(0.028)	(0.024)	(0.004)	
Estimation method	PPML	OLS	OLS	
Oster's bound	NA	0.089	0.015	
N	37909	10706	37909	
R^2	NA	0.036	0.087	
	Learning for new employment (4)	Learning for current job (5)	Learning for cultivation (6)	Learning for other objectives (7)
ln population density	0.015[b]	0.003	0.021[a]	0.015[b]
	(0.006)	(0.005)	(0.007)	(0.008)
Estimation method	OLS	OLS	OLS	OLS
Oster's bound	0.016	0.000	0.019	0.017
N	11290	11290	11145	11290
R^2	0.041	0.048	0.023	0.018

Cluster-robust standard errors in parentheses

^{c} $p < 0.1$, ^{b} $p < 0.05$, ^{a} $p < 0.01$

Table A.15: Learning frequency and purposes (male sample)

Note: In Column 1, I estimate by the Pseudo Poisson Maximum Likelihood (PPML, Santos Silva and Tenreyro 2006) the constant-elasticity regression model (2) relating number of days spent for learning to the municipal population density. Column 2 takes as the outcome variable the log number of days spent for learning in the linear regression model (1). The remaining columns use the linear probability model as the outcome variable is a dummy. Column 6 has slightly fewer observations than columns 4, 5, and 7 because some observations are coded as "unknown." In all specifications, I include as control variables sex, age and its square, marital status, and educational attainment. I calculate Oster's (2019) beta for each regression. Based on a recommendation of Oster (2019), I set the maximal R^2 to 1.3 times the R^2 of each regression. I also assume that observed and unobserved factors are equally related to population density (i.e., the coefficient of proportionality $\delta = 1$ in Oster's notation). Due to its availability, I do not report it when using the PPML. I cluster standard errors at the municipal level.

	English	Other foreign language	Computers etc	Commerce and Business	Elderly care
	(1)	(2)	(3)	(4)	(5)
ln population density	0.020[a]	0.008[a]	0.013[a]	0.014[a]	-0.002[b]
	(0.003)	(0.002)	(0.003)	(0.004)	(0.001)
Estimation method	OLS	OLS	OLS	OLS	OLS
Oster's bound	0.015	0.006	0.009	0.010	-0.003
N	37909	37909	37909	37909	37909
R^2	0.096	0.018	0.028	0.038	0.008
	Housekeeping	Humanities and sciences	Arts and culture	Other	
	(6)	(7)	(8)	(9)	
ln population density	0.008[a]	0.005[b]	0.015	0.002	
	(0.003)	(0.003)	(0.003)	(0.003)	
Estimation method	OLS	OLS	OLS	OLS	
Oster's bound	0.007	0.002	0.012	0.001	
N	37909	37909	37909	37909	
R^2	0.011	0.054	0.032	0.006	

Cluster-robust standard errors in parentheses

^{c} $p < 0.1$, ^{b} $p < 0.05$, ^{a} $p < 0.01$

Table A.16: Learning by subject (male sample)

Note: I use the linear probability model that regresses the dummy for learning a particular subject on the municipal population density. In all specifications, I include as control variables sex, age and its square, marital status, and educational attainment. I calculate Oster's (2019) beta for each regression. Based on a recommendation of Oster (2019), I set the maximal R^2 to 1.3 times the R^2 of each regression. I also assume that observed and unobserved factors are equally related to population density (i.e., the coefficient of proportionality $\delta = 1$ in Oster's notation). I cluster standard errors at the municipal level.

	Number of days learning (1)	Log number of days learning (2)	Engaged in learning (3)	
ln population density	0.063^{a}	0.012	0.017^{a}	
	(0.022)	(0.020)	(0.003)	
Estimation method	PPML	OLS	OLS	
Oster's bound	NA	0.009	0.011	
N	40818	12807	40818	
R^2	NA	0.006	0.055	
	Learning for new employment (4)	Learning for current job (5)	Learning for cultivation (6)	Learning for other objectives (7)
ln population density	0.001	-0.021^{a}	0.031^{a}	-0.006
	(0.003)	(0.006)	(0.006)	(0.007)
Estimation method	OLS	OLS	OLS	OLS
Oster's bound	0.001	-0.025 2	0.026	-0.003
N	13615	13615	13330	13615
R^2	0.008	0.028	0.037	0.016

Cluster-robust standard errors in parentheses

^{c} $p < 0.1$, ^{b} $p < 0.05$, ^{a} $p < 0.01$

Table A.17: Learning frequency and purposes (female sample)

Note: In Column 1, I estimate by the Pseudo Poisson Maximum Likelihood (PPML, Santos Silva and Tenreyro 2006) the constant-elasticity regression model (2) relating number of days spent for learning to the municipal population density. Column 2 takes as the outcome variable the log number of days spent for learning in the linear regression model (1). The remaining columns use the linear probability model as the outcome variable is a dummy. Column 6 has slightly fewer observations than columns 4, 5, and 7 because some observations are coded as "unknown." In all specifications, I include as control variables sex, age and its square, marital status, and educational attainment. I calculate Oster's (2019) beta for each regression. Based on a recommendation of Oster (2019), I set the maximal R^2 to 1.3 times the R^2 of each regression. I also assume that observed and unobserved factors are equally related to population density (i.e., the coefficient of proportionality $\delta = 1$ in Oster's notation). Due to its availability, I do not report it when using the PPML. I cluster standard errors at the municipal level.

	English	Other foreign language	Computers etc	Commerce and Business	Elderly care
	(1)	(2)	(3)	(4)	(5)
ln population density	0.020[a]	0.006[a]	0.006[a]	0.006[a]	-0.005[a]
	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)
Estimation method	OLS	OLS	OLS	OLS	OLS
Oster's bound	0.015	0.005	0.004	0.004	-0.005
N	40818	40818	40818	40818	40818
R^2	0.063	0.017	0.015	0.015	0.011
	Housekeeping	Humanities and sciences	Arts and culture	Other	
	(6)	(7)	(8)	(9)	
ln population density	0.007[b]	0.003[c]	0.011[a]	-0.001	
	(0.003)	(0.002)	(0.003)	(0.002)	
Estimation method	OLS	OLS	OLS	OLS	
Oster's bound	0.006	0.001	0.008	-0.003	
N	40818	40818	40818	40818	
R^2	0.012	0.033	0.044	0.012	

Cluster-robust standard errors in parentheses

[c] $p < 0.1$, [b] $p < 0.05$, [a] $p < 0.01$

Table A.18: Learning by subject (female sample)

Note: I use the linear probability model that regresses the dummy for learning a particular subject on the municipal population density. In all specifications, I include as control variables sex, age and its square, marital status, and educational attainment. I calculate Oster's (2019) beta for each regression. Based on a recommendation of Oster (2019), I set the maximal R^2 to 1.3 times the R^2 of each regression. I also assume that observed and unobserved factors are equally related to population density (i.e., the coefficient of proportionality $\delta = 1$ in Oster's notation). I cluster standard errors at the municipal level.

	Number of days learning (1)	Log number of days learning (2)	Engaged in learning (3)	
ln population density	0.025	0.033	0.016^{a}	
	(0.022)	(0.022)	(0.005)	
Estimation method	PPML	OLS	OLS	
Oster's bound	NA	0.023	0.011	
N	20130	8840	20130	
R^2	NA	0.026	0.043	
	Learning for new employment (4)	Learning for current job (5)	Learning for cultivation (6)	Learning for other objectives (7)
ln population density	-0.001	-0.008^{b}	0.015^{b}	0.000
	(0.001)	(0.003)	(0.006)	(0.007)
Estimation method	OLS	OLS	OLS	OLS
Oster's bound	-0.008	-0.011	0.015	-0.001
N	9418	9418	9220	9418
R^2	0.119	0.113	0.016	0.010

Cluster-robust standard errors in parentheses

^{c} $p < 0.1$, ^{b} $p < 0.05$, ^{a} $p < 0.01$

Table A.19: Learning frequency and purposes (students)

Note: In Column 1, I estimate by the Pseudo Poisson Maximum Likelihood (PPML, Santos Silva and Tenreyro 2006) the constant-elasticity regression model (2) relating number of days spent for learning to the municipal population density. Column 2 takes as the outcome variable the log number of days spent for learning in the linear regression model (1). The remaining columns use the linear probability model as the outcome variable is a dummy. Column 6 has slightly fewer observations than columns 4, 5, and 7 because some observations are coded as "unknown." In all specifications, I include as control variables sex, age and its square, marital status, and educational attainment. I calculate Oster's (2019) beta for each regression. Based on a recommendation of Oster (2019), I set the maximal R^2 to 1.3 times the R^2 of each regression. I also assume that observed and unobserved factors are equally related to population density (i.e., the coefficient of proportionality $\delta = 1$ in Oster's notation). Due to its availability, I do not report it when using the PPML. I cluster standard errors at the municipal level.

	English	Other foreign language	Computers etc	Commerce and Business	Elderly care
	(1)	(2)	(3)	(4)	(5)
ln population density	0.020[a]	0.004	0.000	-0.003	-0.004[b]
	(0.005)	(0.004)	(0.004)	(0.003)	(0.002)
Estimation method	OLS	OLS	OLS	OLS	OLS
Oster's bound	0.016	0.001	-0.005	-0.005	-0.005
N	20130	20130	20130	20130	20130
R^2	0.031	0.057	0.085	0.053	0.032
	Housekeeping	Humanities and sciences	Arts and culture	Other	
	(6)	(7)	(8)	(9)	
ln population density	0.000	0.008[c]	0.013[a]	0.001	
	(0.004)	(0.004)	(0.004)	(0.004)	
Estimation method	OLS	OLS	OLS	OLS	
Oster's bound	-0.002	0.004	0.012	-0.000	
N	20130	20130	20130	20130	
R^2	0.032	0.049	0.028	0.008	

Cluster-robust standard errors in parentheses
[c] $p < 0.1$, [b] $p < 0.05$, [a] $p < 0.01$

Table A.20: Learning by subject (students)

Note: I use the linear probability model that regresses the dummy for learning a particular subject on the municipal population density. In all specifications, I include as control variables sex, age and its square, marital status, and educational attainment. I calculate Oster's (2019) beta for each regression. Based on a recommendation of Oster (2019), I set the maximal R^2 to 1.3 times the R^2 of each regression. I also assume that observed and unobserved factors are equally related to population density (i.e., the coefficient of proportionality $\delta = 1$ in Oster's notation). I cluster standard errors at the municipal level.

	Number of days learning (1)	Log number of days learning (2)	Engaged in learning (3)	
ln population density	0.067^a	0.046^a	0.011^a	
	(0.016)	(0.016)	(0.003)	
Estimation method	PPML	OLS	OLS	
Oster's bound	NA	0.032	0.002	
N	69702	16276	69702	
R^2	NA	0.026	0.108	
	Learning for new employment (4)	Learning for current job (5)	Learning for cultivation (6)	Learning for other objectives (7)
ln population density	0.002	-0.004	0.011^b	-0.001
	(0.001)	(0.004)	(0.005)	(0.004)
Estimation method	OLS	OLS	OLS	OLS
Oster's bound	0.002	-0.005	0.005	-0.000
N	17616	17616	16904	17616
R^2	0.010	0.081	0.039	0.0091

Cluster-robust standard errors in parentheses

^{c} $p < 0.1$, ^{b} $p < 0.05$, ^{a} $p < 0.01$

Table A.21: Learning frequency and purposes (elderly)

Note: In Column 1, I estimate by the Pseudo Poisson Maximum Likelihood (PPML, Santos Silva and Tenreyro 2006) the constant-elasticity regression model (2) relating number of days spent for learning to the municipal population density. Column 2 takes as the outcome variable the log number of days spent for learning in the linear regression model (1). The remaining columns use the linear probability model as the outcome variable is a dummy. Column 6 has slightly fewer observations than columns 4, 5, and 7 because some observations are coded as "unknown." In all specifications, I include as control variables sex, age and its square, marital status, and educational attainment. I calculate Oster's (2019) beta for each regression. Based on a recommendation of Oster (2019), I set the maximal R^2 to 1.3 times the R^2 of each regression. I also assume that observed and unobserved factors are equally related to population density (i.e., the coefficient of proportionality $\delta = 1$ in Oster's notation). Due to its availability, I do not report it when using the PPML. I cluster standard errors at the municipal level.

	English	Other foreign language	Computers etc	Commerce and Business	Elderly care
	(1)	(2)	(3)	(4)	(5)
ln population density	0.006[a]	0.002[a]	0.006[a]	0.002[b]	-0.000
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)
Estimation method	OLS	OLS	OLS	OLS	OLS
Oster's bound	0.003	0.002	0.002	0.001	-0.001
N	69702	69702	69702	69702	69702
R^2	0.062	0.023	0.067	0.033	0.012
	Housekeeping	Humanities and sciences	Arts and culture	Other	
	(6)	(7)	(8)	(9)	
ln population density	0.001	0.002	0.006	0.001	
	(0.001)	(0.001)	(0.002)	(0.001)	
Estimation method	OLS	OLS	OLS	OLS	
Oster's bound	-0.001	-0.001	0.001	-0.001	
N	69702	69702	69702	69702	
R^2	0.040	0.080	0.070	0.019	

Cluster-robust standard errors in parentheses

^{c} $p < 0.1$, ^{b} $p < 0.05$, ^{a} $p < 0.01$

Table A.22: Learning by subject (elderly)

Note: I use the linear probability model that regresses the dummy for learning a particular subject on the municipal population density. In all specifications, I include as control variables sex, age and its square, marital status, and educational attainment. I calculate Oster's (2019) beta for each regression. Based on a recommendation of Oster (2019), I set the maximal R^2 to 1.3 times the R^2 of each regression. I also assume that observed and unobserved factors are equally related to population density (i.e., the coefficient of proportionality $\delta = 1$ in Oster's notation). I cluster standard errors at the municipal level.

	Number of days learning (1)	Log number of days learning (2)	Engaged in learning (3)	
ln population density	0.096^a	0.056^a	0.019^a	
	(0.018)	(0.017)	(0.003)	
ln population density	0.033	-0.020	0.016^c	
× Young dummy	(0.047)	(0.042)	(0.009)	
Estimation method	PPML	OLS	OLS	
N	78727	23513	78727	
R^2	NA	0.014	0.069	
	Learning for new employment (4)	Learning for current job (5)	Learning for cultivation (6)	Learning for other objectives (7)
ln population density	0.010^b	-0.008^c	0.026^a	0.001
	(0.004)	(0.005)	(0.005)	(0.004)
ln population density	-0.007	-0.006	-0.000	0.018
× Young dummy	(0.009)	(0.014)	(0.014)	(0.015)
Estimation method	OLS	OLS	OLS	OLS
N	24905	24905	24475	24905
R^2	0.017	0.070	0.032	0.025

Cluster-robust standard errors in parentheses

^{c} $p < 0.1$, ^{b} $p < 0.05$, ^{a} $p < 0.01$

Table A.23: Learning frequency and purposes (interaction with young dummy)

Note: In Column 1, I estimate by the Pseudo Poisson Maximum Likelihood (PPML, Santos Silva and Tenreyro 2006) the constant-elasticity regression model (2) relating number of days spent for learning to the municipal population density. Column 2 takes as the outcome variable the log number of days spent for learning in the linear regression model (1). The remaining columns use the linear probability model as the outcome variable is a dummy. Column 6 has slightly fewer observations than columns 4, 5, and 7 because some observations are coded as "unknown." In all specifications, I include as control variables sex, age and its square, marital status, and educational attainment. I calculate Oster's (2019) beta for each regression. Based on a recommendation of Oster (2019), I set the maximal R^2 to 1.3 times the R^2 of each regression. I also assume that observed and unobserved factors are equally related to population density (i.e., the coefficient of proportionality $\delta = 1$ in Oster's notation). Due to its availability, I do not report it when using the PPML. I cluster standard errors at the municipal level.

	English	Other foreign language	Computers etc	Commerce and Business	Elderly care
	(1)	(2)	(3)	(4)	(5)
ln population density	0.019[a]	0.007[a]	0.007[b]	0.009[b]	-0.003[a]
	(0.002)	(0.001)	(0.002)	(0.002)	(0.001)
ln population density	0.009	-0.001	0.019[a]	0.009	-0.004
× Young dummy	(0.005)	(0.004)	(0.006)	(0.008)	(0.002)
Estimation method	OLS	OLS	OLS	OLS	OLS
N	78727	78727	78727	78727	78727
R^2	0.080	0.017	0.028	0.039	0.013
	Housekeeping	Humanities and sciences	Arts and culture	Other	
	(6)	(7)	(8)	(9)	
ln population density	0.005[a]	0.004[a]	0.011[a]	0.001	
	(0.002)	(0.002)	(0.002)	(0.002)	
ln population density	0.019[c]	-0.003	0.015[c]	-0.007	
× Young dummy	(0.011)	(0.006)	(0.009)	(0.007)	
Estimation method	OLS	OLS	OLS	OLS	
N	78727	78727	78727	78727	
R^2	0.037	0.050	0.038	0.007	

Cluster-robust standard errors in parentheses
[c] $p < 0.1$, [b] $p < 0.05$, [a] $p < 0.01$

Table A.24: Learning by subject (interaction with young dummy)

Note: I use the linear probability model that regresses the dummy for learning a particular subject on the municipal population density. In all specifications, I include as control variables sex, age and its square, marital status, and educational attainment. I calculate Oster's (2019) beta for each regression. Based on a recommendation of Oster (2019), I set the maximal R^2 to 1.3 times the R^2 of each regression. I also assume that observed and unobserved factors are equally related to population density (i.e., the coefficient of proportionality $\delta = 1$ in Oster's notation). I cluster standard errors at the municipal level.

	Number of days learning	Log number of days learning	Engaged in learning	
	(1)	(2)	(3)	
ln population density	0.114^a	0.046^a	0.021^a	
	(0.020)	(0.017)	(0.003)	
ln population density	-0.046	0.030	0.003	
× High skill	(0.039)	(0.030)	(0.008)	
Estimation method	PPML	OLS	OLS	
N	78727	23513	78727	
R^2	NA	0.015	0.077	
	Learning for new employment	Learning for current job	Learning for cultivation	Learning for other objectives
	(4)	(5)	(6)	(7)
ln population density	0.007	-0.014^a	0.028^a	0.002
	(0.005)	(0.005)	(0.005)	(0.005)
ln population density	0.006	0.029^a	-0.007	0.008
× High skill	(0.009)	(0.009)	(0.010)	(0.009)
Estimation method	OLS	OLS	OLS	OLS
N	24905	24905	24475	24905
R^2	0.019	0.100	0.032	0.026

Cluster-robust standard errors in parentheses

^{c} $p < 0.1$, ^{b} $p < 0.05$, ^{a} $p < 0.01$

Table A.25: Learning frequency and purposes (interaction with high skill job)

Note: In Column 1, I estimate by the Pseudo Poisson Maximum Likelihood (PPML, Santos Silva and Tenreyro 2006) the constant-elasticity regression model (2) relating number of days spent for learning to the municipal population density. Column 2 takes as the outcome variable the log number of days spent for learning in the linear regression model (1). The remaining columns use the linear probability model as the outcome variable is a dummy. Column 6 has slightly fewer observations than columns 4, 5, and 7 because some observations are coded as "unknown." In all specifications, I include as control variables sex, age and its square, marital status, and educational attainment. I calculate Oster's (2019) beta for each regression. Based on a recommendation of Oster (2019), I set the maximal R^2 to 1.3 times the R^2 of each regression. I also assume that observed and unobserved factors are equally related to population density (i.e., the coefficient of proportionality $\delta = 1$ in Oster's notation). Due to its availability, I do not report it when using the PPML. I cluster standard errors at the municipal level.

	English	Other foreign language	Computers etc	Commerce and Business	Elderly care
	(1)	(2)	(3)	(4)	(5)
ln population density	0.019[a]	0.007[a]	0.008[a]	0.009[a]	-0.001[c]
	(0.002)	(0.001)	(0.002)	(0.002)	(0.001)
ln population density × High skill	0.008	0.000	0.015[b]	0.007	-0.014[a]
	(0.007)	(0.003)	(0.007)	(0.006)	(0.003)
Estimation method	OLS	OLS	OLS	OLS	OLS
N	78727	78727	78727	78727	78727
R^2	0.083	0.017	0.033	0.039	0.018
	Housekeeping	Humanities and sciences	Arts and culture	Other	
	(6)	(7)	(8)	(9)	
ln population density	0.008[a]	0.004[b]	0.013[a]	0.001	
	(0.002)	(0.002)	(0.002)	(0.002)	
ln population density × High skill	-0.003	0.002	0.007	-0.002	
	(0.005)	(0.007)	(0.006)	(0.007)	
Estimation method	OLS	OLS	OLS	OLS	
N	78727	78727	78727	78727	
R^2	0.036	0.055	0.043	0.011	

Cluster-robust standard errors in parentheses

^{c} $p < 0.1$, ^{b} $p < 0.05$, ^{a} $p < 0.01$

Table A.26: Learning by subject (interaction with high skill job)

Note: I use the linear probability model that regresses the dummy for learning a particular subject on the municipal population density. In all specifications, I include as control variables sex, age and its square, marital status, and educational attainment. I calculate Oster's (2019) beta for each regression. Based on a recommendation of Oster (2019), I set the maximal R^2 to 1.3 times the R^2 of each regression. I also assume that observed and unobserved factors are equally related to population density (i.e., the coefficient of proportionality $\delta = 1$ in Oster's notation). I cluster standard errors at the municipal level.

B Additional results

How people learn. Table B.1 presents the results on learning methods.[B.1] Although not conclusive, there appears to be a tendency that in denser cities, people more frequently utilize education-market-based methods that require physical presence (e.g., private classes and vocational schools). In contrast, people in less dense cities tend to rely less on non-education-market-based methods, such as publicly provided learning opportunities or workplace training outside of working hours.[B.2] This aligns with the view that cities provide more learning opportunities, as private education providers can benefit from agglomeration economies, which tend to be strong in the service sector (Glaeser, Kolko and Saiz 2001; Morikawa 2011; Leonardi and Moretti 2023).[B.3] Consistent with this, Appendix F provides additional evidence that private-sector education opportunities are more abundant in denser areas.

I have argued that the private market provides more learning opportunities in denser cities, thereby lowering the cost of engaging in learning activities. At the same time, this paper has documented evidence that labor market demand for skills, which rises with population density, serves as a crucial incentive for off-the-job learning. To assess the relative importance of these factors in shaping learning behavior, I additionally control for the density of private education establishments as a proxy for the availability of private learning opportunities.[B.4]

Tables B.3 and B.4 replicate the analyses of Tables 1 and 2, now including controls for the density of private education establishments. Overall, Table B.3 shows that population density remains a significant predictor of most outcomes related to learning frequency and purpose. In contrast, the density of education establishments is significant for some outcomes but generally exhibits weaker significance. Similarly, Table B.4 suggests that the coefficients on log population density remain largely consistent with the main estimates presented in Table 2. That said, I acknowledge that the results in Tables B.3 and B.4 are somewhat noisy due to the high correlation between population density and the density of education facilities ($\rho \simeq 0.92$).[B.5]

[B.1] For the same reasons as in the analysis of learning purposes, I focus on individuals who engaged in some form of learning.

[B.2] Greater reliance on publicly provided opportunities suggests that the public sector may step in to compensate for the lack of private learning options.

[B.3] In addition to supply-side factors, the chosen learning method also depends on the subject of study. For example, elderly care is often learned through lecture-based meetings, while foreign languages are typically learned via TV and radio. These factors may contribute to the statistically significant estimates in Table B.1.

[B.4] The density measure is taken as per habitable area in 2015, consistent with population density. I use data from the 2016 Economic Census, accessed via the Regional Economy Society Analyzing System (https://resas.go.jp/, in Japanese, last accessed on June 21, 2024). For the definition of education establishments that are not formal educational institutions, see https://www.soumu.go.jp/main_content/000290734.pdf (in Japanese). Unfortunately, the data on the number of education establishments by category (e.g., English schools, computer schools) are unavailable.

[B.5] This highlights the difficulty of disentangling demand from supply. The results in Tables B.3 and B.4 suggest that, in two cities with the same supply of private education, workers will engage in more off-the-job learning in the denser one. While this suggests the importance of labor market demand for learning in cities, caution is required in this interpretation, as the equilibrium supply of private education is itself shaped by demand. To fully isolate the effect of labor market demand, an instrumental variable that exogenously shifts the supply of education would be necessary. I leave this important issue for future research.

Learning with whom. To examine whether learning behavior in denser cities is more likely to involve social interactions—which could be relevant for knowledge spillovers—I analyze responses from the JTUS time-use diary. Specifically, I regress population density on four dummy variables indicating whether a person engages in learning (i) alone, (ii) with family members, (iii) with colleagues or classmates, or (iv) with others. Table B.2 shows that people in denser cities are more likely to engage in learning alone or with family members. However, the likelihood of learning with colleagues or other people shows little association with population density. Thus, in the context of off-the-job learning, greater knowledge spillovers through interactions with new people in denser cities do not appear to be present. It is important to note, however, that this does not rule out the broader significance of interpersonal interactions and knowledge spillovers in urban environments. Such mechanisms may operate through other dimensions of learning, such as workplace communication or professional networking (e.g., Charlot and Duranton 2004; Sandvik, Saouma, Seegert and Stanton 2020).

	Classes (public) (1)	Classes (private) (2)	Classes (universities etc) (3)	Lecture meetings (4)	Home-study courses (5)
ln population density	-0.009^a	0.006^b	-0.001	-0.009^a	-0.003
	(0.003)	(0.003)	(0.001)	(0.002)	(0.002)
Estimation method	OLS	OLS	OLS	OLS	OLS
Oster's bound	-0.009	0.005	-0.002	-0.010	-0.004
N	24304	24905	24905	24905	24905
R^2	0.019	0.021	0.015	0.019	0.007
	TV and radio (6)	Workplace (outside working hours) (7)	Vocational schools etc (8)	Training by employment service center (9)	Other (10)
ln population density	0.008^b	-0.007^c	0.005^a	-0.001^b	0.010^b
	(0.004)	(0.004)	(0.002)	(0.001)	(0.004)
Estimation method	OLS	OLS	OLS	OLS	OLS
Oster's bound	0.008	-0.008	0.005	-0.001	0.009
N	24905	24905	24905	24905	24905
R^2	0.026	0.034	0.009	0.004	0.007

Cluster-robust standard errors in parentheses
^{c} $p < 0.1$, ^{b} $p < 0.05$, ^{a} $p < 0.01$

Table B.1: Learning by method

Note: I use the linear probability model that regresses the dummy for learning in a particular method on the municipal population density. The dummy of learning by classes (public) has slightly fewer observations than the other dummy variables on methods because some observations are coded as "unknown" in the original data. In all specifications, I include as control variables sex, age and its square, marital status, and educational attainment. I calculate Oster's (2019) beta for each regression. Based on a recommendation of Oster (2019), I set the maximal R^2 to 1.3 times the R^2 of each regression. I also assume that observed and unobserved factors are equally related to population density (i.e., the coefficient of proportionality $\delta = 1$ in Oster's notation). I cluster standard errors at the municipal level.

	Alone (1)	Family members (2)	Colleagues or classmates (3)	Others (4)
ln population density	0.002^b	0.002^b	0.0002	0.0003
	(0.001)	(0.001)	(0.0004)	(0.0003)
N	154105	154105	154105	154105
R^2	0.010	0.006	0.003	0.003

Cluster-robust standard errors in parentheses
^{c} $p < 0.1$, ^{b} $p < 0.05$, ^{a} $p < 0.01$

Table B.2: Learning with whom

Note: In Column 1, I regress a dummy that takes one if the person engages in some learning alone. Similarly, Column 2 uses a dummy that takes one if the person engages in some learning with family member, Column 3 uses a dummy that takes one if the person engages in some learning with colleagues or classmates, and Column 4 uses a dummy that takes one if the person engages in some learning with other people. The sample size is larger than the main text because one is required to report his or her behavior for two consecutive days. In all specifications, I include as control variables sex, age and its square, marital status, and educational attainment. I also control for the survey day of week since this analysis uses the time diary question about a particular day. I cluster standard errors at the municipal level.

	Number of days learning (1)	Log number of days learning (2)	Engaged in learning (3)	
ln population density	0.045^b	0.025	0.010^b	
	(0.022)	(0.022)	(0.005)	
Density of education establishments	0.013^a	0.006	0.003^b	
	(0.004)	(0.004)	(0.001)	
Estimation method	PPML	OLS	OLS	
N	78727	23513	78727	
R^2	NA	0.015	0.030	
	Learning for new employment (4)	Learning for current job (5)	Learning for cultivation (6)	Learning for other objectives (7)
ln population density	0.011^a	-0.025^a	0.024^a	-0.000
	(0.004)	(0.009)	(0.007)	(0.007)
Density of education establishments	-0.001	0.004^c	0.000	0.001
	(0.001)	(0.002)	(0.002)	(0.001)
Estimation method	OLS	OLS	OLS	OLS
N	24905	24905	24475	24905
R^2	0.017	0.071	0.032	0.025

Cluster-robust standard errors in parentheses
^{c} $p < 0.1$, ^{b} $p < 0.05$, ^{a} $p < 0.01$

Table B.3: Learning frequency and purposes (controlling for density of education establishments)

Note: In Column 1, I estimate by the Pseudo Poisson Maximum Likelihood (PPML, Santos Silva and Tenreyro 2006) the constant-elasticity regression model (2) relating number of days spent for learning to the municipal population density. Column 2 takes as the outcome variable the log number of days spent for learning in the linear regression model (1). The remaining columns use the linear probability model as the outcome variable is a dummy. Column 6 has slightly fewer observations than columns 4, 5, and 7 because some observations are coded as "unknown." In all specifications, I include as control variables sex, age and its square, marital status, and educational attainment. I calculate Oster's (2019) beta for each regression. Based on a recommendation of Oster (2019), I set the maximal R^2 to 1.3 times the R^2 of each regression. I also assume that observed and unobserved factors are equally related to population density (i.e., the coefficient of proportionality $\delta = 1$ in Oster's notation). Due to its availability, I do not report it when using the PPML. I cluster standard errors at the municipal level.

	English	Other foreign language	Computers etc	Commerce and Business	Elderly care
	(1)	(2)	(3)	(4)	(5)
ln population density	0.008^a	0.002	0.004	0.007	-0.004^a
	(0.003)	(0.003)	(0.003)	(0.005)	(0.001)
Density of education establishments	0.003^a	0.001^c	0.002^a	0.001	0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)
Estimation method	OLS	OLS	OLS	OLS	OLS
N	78727	78727	78727	78727	78727
R^2	0.081	0.017	0.027	0.039	0.013
	Housekeeping	Humanities and sciences	Arts and culture	Other	
	(6)	(7)	(8)	(9)	
ln population density	0.004	0.000	-0.002	-0.003	
	(0.003)	(0.003)	(0.002)	(0.003)	
Density of education establishments	0.000	0.001	0.004^a	0.001	
	(0.001)	(0.001)	(0.001)	(0.001)	
Estimation method	OLS	OLS	OLS	OLS	
N	78727	78727	78727	78727	
R^2	0.036	0.050	0.041	0.007	

Cluster-robust standard errors in parentheses
^{c} $p < 0.1$, ^{b} $p < 0.05$, ^{a} $p < 0.01$

Table B.4: Learning by subject (controlling for density of education establishments)

Note: I use the linear probability model that regresses the dummy for learning a particular subject on the municipal population density. In all specifications, I include as control variables sex, age and its square, marital status, and educational attainment. I calculate Oster's (2019) beta for each regression. Based on a recommendation of Oster (2019), I set the maximal R^2 to 1.3 times the R^2 of each regression. I also assume that observed and unobserved factors are equally related to population density (i.e., the coefficient of proportionality $\delta = 1$ in Oster's notation). I cluster standard errors at the municipal level.

C Endogeneity of population

While not a problem for a descriptive analysis, identification of causal impact of population density on learning is problematic if some unobserved factors simultaneously affect population and learning. To address these issues, I follow Ciccone and Hall (1996) and use as an IV a long lag of the population density using the 1920 population census, the first modern population census in Japan.

I use the GIS version of the 1930 population census, which also records population from the 1920 census, by Yuji Murayama (http://giswin.geo.tsukuba.ac.jp/teacher/murayama/data.html, in Japanese, last accessed on September 8, 2022). I combine this with the shape file of municipalities as of 2016 and proportionally assign the historical municipal population to current municipalities based on the overlapping area size. Then I divide by the physical area of each municipality to address another concern that the habitable area could be determined endogenously (Hayakawa, Koster, Tabuchi and Thisse 2021). Note that Okinawa prefecture is not covered by this dataset and thus dropped in the IV analysis. This matters little in practice because my OLS results change little by dropping Okinawa prefecture. I use the GMM Poisson regression for estimating equation (2) and 2SLS regression for equation (1).

Tables C.1–C.3 present the IV results. I find that the IV results are fairly similar to the baseline OLS results. The similarity of the IV results implies the limited importance of the endogeneity of population size, as in Ciccone and Hall (1996) and De la Roca and Puga (2017).

	Number of days learning	ln number of days learning	Engaged in learning	
	(1)	(2)	(3)	
ln population density	0.130^a	0.080^a	0.025^a	
	(0.027)	(0.025)	(0.005)	
Estimation method	GMM	2SLS	2SLS	
N	77236	23084	77236	
R^2	NA	0.014	0.069	
	Learning for new employment	Learning for current job	Learning for cultivation	Learning for other objectives
	(4)	(5)	(6)	(7)
ln population density	0.011	-0.006	0.024^a	0.021^b
	(0.008)	(0.008)	(0.007)	(0.009)
Estimation method	2SLS	2SLS	2SLS	2SLS
N	24438	24438	24013	24438
R^2	0.016	0.070	0.032	0.025

Cluster-robust standard errors in parentheses

^{c} $p < 0.1$, ^{b} $p < 0.05$, ^{a} $p < 0.01$

Table C.1: Learning frequency and purposes (IV results)

Note: In Column 1, I estimate by the GMM constant-elasticity model (2) regressing number of days spent for learning on the municipal population density. In Column 2, I estimate by two-stage-least-square (2SLS) the linear regression model taking log population density as the outcome variable. In the remaining columns, I estimate the linear probability model. In all columns, I use as an IV for the population density in 2015 the population density from 1920 population census. The first-stage F-statistic is 43.85 for Column 1 and 3, 32.97 for Column 2, 34.22 for Column 6 and 33.53 for Columns 4, 5, and 7. Column 6 has slightly fewer observations than Columns 4, 5, and 7 because some observations are coded as “unknown." In all specifications, I include as control variables sex, age and its square, marital status, and educational attainment. I cluster standard errors at the municipal level.

	English	Other foreign language	Computers etc	Commerce and Business	Elderly care
	(1)	(2)	(3)	(4)	(5)
ln population density	0.023[a]	0.010[a]	0.013[a]	0.011[a]	-0.004[a]
	(0.004)	(0.002)	(0.003)	(0.004)	(0.001)
Estimation method	2SLS	2SLS	2SLS	2SLS	2SLS
N	77236	77236	77236	77236	77236
R^2	0.080	0.016	0.027	0.039	0.013
	Housekeeping	Humanities and sciences	Arts and culture	Other	
	(6)	(7)	(8)	(9)	
ln population density	0.012[a]	0.004	0.019[a]	0.002	
	(0.004)	(0.003)	(0.004)	(0.003)	
Estimation method	2SLS	2SLS	2SLS	2SLS	
N	77236	77236	77236	77326	
R^2	0.036	0.050	0.037	0.007	

Cluster-robust standard errors in parentheses

^{c} $p < 0.1$, ^{b} $p < 0.05$, ^{a} $p < 0.01$

Table C.2: Learning by subject (IV results)

Note: I use the linear probability model that regresses the dummy for learning a particular subject on the municipal population density. I use as an IV for the population density in 2015 the population density from 1920 population census. The first-stage F-statistic is 43.85. In all specifications, I include as control variables sex, age and its square, marital status, and educational attainment. I cluster standard errors at the municipal level.

	Classes (public)	Classes (private)	Classes (universities etc)	Lecture meetings	Home-study courses
	(1)	(2)	(3)	(4)	(5)
ln population density	-0.011[a]	0.013[b]	-0.003[c]	-0.007	-0.000
	(0.002)	(0.006)	(0.002)	(0.005)	(0.004)
Estimation method	2SLS	2SLS	2SLS	2SLS	2SLS
N	23847	24438	24438	24438	24438
R^2	0.019	0.021	0.014	0.019	0.007
	TV and radio	Workplace (outside working hours)	Vocational schools etc	Training by employment service center	Other
	(6)	(7)	(8)	(9)	(10)
ln population density	0.012[b]	-0.001	0.004	-0.001	0.005
	(0.006)	(0.006)	(0.003)	(0.001)	(0.009)
Estimation method	2SLS	2SLS	2SLS	2SLS	2SLS
N	24438	24438	24438	24438	24438
R^2	0.021	0.034	0.009	0.004	0.001

Cluster-robust standard errors in parentheses

[c] $p < 0.1$, [b] $p < 0.05$, [a] $p < 0.01$

Table C.3: Learning by method (IV results)

Note: I use the linear probability model that regresses the dummy for learning in a particular method on the municipal population density. I use as an IV for the population density in 2015 the population density from 1920 population census. The first-stage F-statistic is 34.47 for Column 1 and 33.53 for the other columns. The first column has slightly fewer observations because some observations are coded as “unknown." In all specifications, I include as control variables sex, age and its square, marital status, and educational attainment. I cluster standard errors at the municipal level.

D Sorting and non-movers

Since my data is cross-sectional, the correlation between learning outcomes and population density might be driven by sorting based on the unobserved heterogeneity of residents, such as a high return from learning and an intrinsic motivation. The urban wage premium literature has addressed this issue by including individual fixed effects (e.g., Glaeser and Maré 2001; De la Roca and Puga 2017). Unfortunately, I cannot use the same strategy in this paper because the JTUS has no panel structure.[D.1]

Instead, I follow Charlot and Duranton (2004) and focus on non-movers to assess how endogenous sorting might affect my results. Since the 2016 JTUS data do not contain questions about the previous residence, I investigate the 1986 JTUS data containing such information. Although in a different format than the 2016 JTUS, the 1986 JTUS asks frequency, subject, and method of learning in a different format than the 2016 JTUS.[D.2] Unfortunately, purpose of learning was not asked in the 1986 survey. I analyze the association between population density and learning outcomes by investigating the full sample and the subsample of (i) people living in the same residence for more than four years and (ii) people staying in the prefecture in which they went through compulsory education, which ends at the age of 15 or 16 in Japan. I use the same set of controls as the 2016 data.

Table D.1 reports the results. I find a similar effect size of population density for the full sample and the subsample of non-movers. In particular, for estimates that are statistically significant in the full sample, focusing on the non-mover samples never changes the qualitative result. This limited importance of geographical mobility for my results is consistent with evidence that the migration decision is not necessarily determined by the expected returns, but by other factors such as self-confidence (De la Roca, Ottaviano and Puga forthcoming). Overall, while not conclusive, this result suggests that my results on learning frequency, subjects, and methods might not be driven by residential sorting based on unobserved characteristics.

[D.1] Note, however, that including individual fixed effects could exacerbate the bias because the identification relies only on migrants, whose migration choices are made endogeneously (De la Roca and Puga 2017). In particular, the sorting based on the magnitude of β is not addressed by the individual fixed effects (Bacolod, Blum, Rangel and Strange 2023).

[D.2] The 2016 JTUS first asks learning frequency for each *skill* as seen in Figure 1. In contrast, the 1986 JTUS first asks learning frequency for each *method*. Total number of days spent for learning is now defined by summing them up. For each method, it then asks the subjects that were learned by this method. The listed methods and skills also somewhat differ from the 2016 JTUS. See https://d-infra.ier.hit-u.ac.jp/Japanese/statistical-yb/b007.html for the questionnaire of the 1986 JTUS data (in Japanese, last accessed on October 2, 2022).

	Number of days learning (1)	Log number of days learning (2)	Engaged in learning (3)	Foreign languages (4)	Business and commerce (5)
Full sample	0.091^{a}	0.155^{a}	-0.001	0.010^{a}	0.007^{a}
	(0.014)	(0.010)	(0.003)	(0.001)	(0.001)
Staying in the prefecture	0.089^{a}	0.161^{a}	-0.003	0.008^{a}	0.006^{a}
of compulsory education	(0.016)	(0.012)	(0.001)	(0.001)	(0.001)
Not moving	0.117^{a}	0.173^{a}	0.003	0.008^{a}	0.007^{a}
in four years	(0.016)	(0.011)	(0.004)	(0.001)	(0.001)
	Engineering (1)	Medicine (2)	Cooking and beauty (3)	Housekeeping (4)	Childbearing (5)
Full sample	0.000	-0.002^{a}	-0.001^{c}	-0.000	-0.004^{a}
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Staying in the prefecture	0.001	-0.003^{a}	-0.002^{a}	-0.000	-0.004^{a}
of compulsory education	(0.001)	(0.001)	(0.001)	(0.005)	(0.001)
Not moving	0.001	-0.002^{a}	-0.002^{a}	0.000	-0.004^{a}
in four years	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
	Education and social welfare (1)	Social science (2)	Natural and science (3)	Arts and culture (4)	Current affairs (5)
Full sample	-0.011^{a}	-0.000	-0.003^{a}	0.004^{a}	-0.002^{a}
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Staying in the prefecture	-0.011^{a}	-0.000	-0.004^{a}	0.005^{a}	-0.002^{c}
of compulsory education	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Not moving	-0.009^{a}	0.001	-0.004^{a}	0.005^{a}	-0.002^{b}
in four years	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
	Miscellaneous school (1)	Vocational school (2)	Classes (public) (3)	Classes (private) (4)	Lecture meetings (5)
Full sample	0.002^{a}	-0.0004^{b}	-0.010^{a}	0.006^{a}	-0.011^{a}
	(0.0003)	(0.0002)	(0.001)	(0.001)	(0.001)
Staying in the prefecture	0.001^{a}	-0.0003	-0.009^{a}	0.005^{a}	-0.011^{a}
of compulsory education	(0.0003)	(0.0003)	(0.001)	(0.005)	(0.002)
Not moving	0.001^{a}	-0.0004^{c}	-0.010^{a}	0.007^{a}	-0.010^{a}
in four years	(0.0003)	(0.0002)	(0.001)	(0.001)	(0.002)
	Correspondence study (1)	TV and radio (2)	Workplace (outside (working hours) (3)	Group study (4)	Study alone (5)
Full sample	0.001^{b}	0.007^{a}	-0.001	-0.004^{a}	0.007^{a}
	(0.0005)	(0.001)	(0.001)	(0.001)	(0.002)
Staying in the prefecture	0.000	0.006^{a}	-0.003^{a}	-0.004^{a}	0.005^{a}
of compulsory education	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Not moving	0.0013^{a}	0.008^{a}	-0.000	-0.003^{a}	0.007^{a}
in four years	(0.0005)	(0.002)	(0.001)	(0.001)	(0.002)

Cluster-robust standard errors in parentheses

^{c} $p < 0.1$, ^{b} $p < 0.05$, ^{a} $p < 0.01$

Table D.1: Learning and population density among non-movers (1986 JTUS)

Note: Using the 1986 JTUS data, I estimate the association between population density and learning outcomes using the linear regression, except for Column 1 in which I use the constant-elasticity model (2). Column 1 is estimated by the PPML while the remaining columns are estimated by the OLS. Coefficients of the log population density is reported, separately for the full sample ($N = 149,872$), the sample of those staying in the prefecture of compulsory education ($N = 114,595$), and the sample of those staying in the same residence over the past four years ($N = 113,307$). In all specifications, I include as control variables sex, age and its square, marital status, and educational attainment. Standard errors are clustered at the municipal level.

E Replicating the urban earnings premium

I examine the relationship between earnings and population density to replicate the urban earnings premium in my JTUS data. Although it has already been shown that larger cities have higher income in Japan (e.g., Tabuchi and Yoshida 2000), my analysis would offer some new insights by using more recent individual-level microdata.[E.1]

I create a continuous measure of the individual income from the JTUS. For those with positive earnings, the JTUS 2016 asks the individual annual earnings in 15 categories. I assign the median value of each category to create a continuous measure of the annual earnings.[E.2] I then regress the log individual annual earnings on the log population density and the same set of control variables as Section 3 (i.e., sex, age and its square, marital status, and education attainment). Column 1 of Table E.1 shows that 1% increase in density is associated with 0.039% increase in income. The magnitude of this estimate is consistent with other estimates and is close to 0.04% in the meta-analysis of the urban wage premium by Ahlfeldt and Pietrostefani (2019). Higher earnings in denser cities are also consistent with the view that more frequent learning in cities documented in this paper actually improves productivity.

Note that this OLS regression excludes those with no income. To address this issue, I also estimate the constant-elasticity regression model (2). Column 2 of Table E.1 shows a somewhat larger urban earnings premium of 0.058%, but broadly consistent with the prior studies (e.g., De la Roca and Puga 2017; Ahlfeldt and Pietrostefani 2019).

Strictly speaking, earnings and wages may differ due to differences in working hours. In particular, earnings are mechanically higher if workers in denser cities work more, as documented by Rosenthal and Strange (2008) for professional workers in the US. However, using the time-use diary question of the JTUS, I find that denser cities do not have longer working hours or, if any, have somewhat shorter working hours, which is consistent with the result of Rosenthal and Strange (2008) for non-professional workers in the US.[E.3] This implies that in my context, the urban earnings premium would be similar to or somewhat smaller than the urban wage premium.

[E.1] Tabuchi and Yoshida (2000) conducts the city-level analysis and does not use individual-level microdata. In general, evidence on urban earnings premium is scant in the Japanese context (see Higashi 2022 for a recent survey). Another subtle difference of my analysis from Tabuchi and Yoshida (2000) is that I measure urbanity by population density, while Tabuchi and Yoshida (2000) use the size of the city. See also Ahlfeldt and Pietrostefani (2019) on this issue.

[E.2] 15 categories are (1) 0–0.5 million yen, (2) 0.5–1 million yen, (3) 1–1.5 million yen, (4) 1.5–2 million yen, (5) 2–2.5 million yen, (6) 2.5–3 million yen, (7) 3–4 million yen, (8) 4–5 million yen, (9) 5–6 million yen, (10) 6–7 million yen, (11) 7–8 million yen, (12) 8–9 million yen, (13) 9–10 million yen, (14) 10–15 million yen, (15) more than 15 million yen. Category 15 does not have an upper bound, so I assign 15 million yen to everyone in this category.

[E.3] For example, the constant-elasticity regression model (2), regressing working hours on the log population density, yields the coefficient of -0.024 (S.E., 0.004).

	ln individual income	
	(1)	(2)
ln population density	0.039^a	0.058^a
	(0.006)	(0.006)
Estimation method	OLS	PPML
N	66545	67057

Cluster-robust standard errors in parentheses
^{c} $p < 0.1$, ^{b} $p < 0.05$, ^{a} $p < 0.01$

Table E.1: Urban earnings premium in the JTUS data

Note: Column 1 presents the linear regression result of log individual income on log population density. Column 2 adopts the constant-elasticity model in equation (2) to include those with zero income. The estimation is by PPML recommended by Santos Silva and Tenreyro (2006). In all specifications, I include as control variables sex, age and its square, marital status, and educational attainment. I cluster standard errors at the municipal level.

F Supplementary evidence from additional datasets

My paper uses the Japanese Time-Use Survey (JTUS) to investigate learning behavior. This section provides supplementary analysis to my arguments using additional datasets. Unlike the JTUS data that are at the individual level, the unit of observations for all of the following analyses is at the municipality level.

More skilled workforce in denser areas. In Section 3.1, I document that people engage in more learning behavior in denser cities, and I argue that this is consistent with the specialization of cities in skill-intensive jobs. To check this claim, I analyze whether denser cities have more skilled workers. To do this, I use the 2015 Population Census to analyze the share of workers engaging in professional jobs, such as researchers and medical specialists.[F.1] By regressing the share of professional workers on log population density, I find that 1% increase in population density is associated with 0.15% increase in the share (S.E. 0.005). That is, denser areas tend to have more skilled workers, which is consistent with higher skilled demand in denser areas.

More cultural activities in denser areas. In Section 3.2, I find that people in denser cities tend to engage in learning for cultivating themselves. This finding is consistent with more cultural activities in denser areas. To further support this, I use the 2015 Population Census to analyze the share of workers engaging in cultural jobs, such as writers and musicians.[F.2] By regressing the share of cultural workers on log population density, I find that 1% increase in population density is associated with 0.08% increase in the share (S.E. 0.005). That is, denser areas tend to have more cultural workers, which is consistent with more cultural activities in denser areas.

Higher demand for elderly care in less dense areas. In Section 3.3, I find that people in denser cities tend to learn any subject, but the opposite is true for elderly care skills. I then interpret this as suggestive evidence that local skill demand shapes the subjects of learning as elderly care skills are highly demanded in rural areas of Japan, where population aging is severe. To supplement this argument, I provide two pieces of suggestive evidence that rural areas are likely to have high demand for elderly care. First, using the 2015 Population Census, I regress the share of people over the age of 65 on log population density. I find that 1% increase in population density is associated with 2.88% decrease in the share (S.E. 0.009). That is, less dense areas tend to have more older population, which then implies higher demand for elderly care. Second, I analyze the per-capita public spending for elderly welfare.[F.3] Since many elderly care services are subsidized or publicly-funded, it is likely to serve as a good proxy of the local demand of the elderly care. By regressing the per-capita spending for elderly welfare on log population density, I find that

[F.1] See https://www.e-stat.go.jp/koumoku/koumoku_teigi/F for more detailed definition of professional jobs (in Japanese, last accessed on June 21, 2024).

[F.2] See https://www.e-stat.go.jp/koumoku/koumoku_teigi/F for more detailed definition of professional jobs (in Japanese, last accessed on June 21, 2024).

[F.3] The data is taken from municipal fiscal data. See https://www.e-stat.go.jp/koumoku/koumoku_teigi/D for more details (in Japanese, last accessed on June 21, 2024).

1% increase in population density is associated with 0.24% decrease in the per-capita spending for the elderly (S.E. 0.006). Overall, both results suggest that rural areas have higher demand for elderly care services.

More learning opportunities in denser areas. In Appendix B, I find that people in denser areas tend to more often utilize education-market-based methods requiring physical presence (i.e., private classes and vocational schools etc). I then argue that this result is consistent with more availability of such learning opportunities due to agglomeration economies. To check this, I analyze the density of private education establishments (excluding formal educational institutes).[F.4] Regressing the log density of education establishments on the log population density, I find that 1% increase in population density is associated with 0.29% increase in the number of education establishments (S.E. 0.03). Therefore, denser areas are indeed characterized by more availability of market-based education opportunities.

[F.4] I take the 2016 Economic Census data from Regional Economy Society Analyzing System (https://resas.go.jp/, in Japanese, last accessed on June 21, 2024). For definition of education establishments that are not formal educational institutes, see https://www.soumu.go.jp/main_content/000290734.pdf (in Japanese).

References for the Appendix

Ahlfeldt, Gabriel M and Elisabetta Pietrostefani, "The economic effects of density: A synthesis," *Journal of Urban Economics*, 2019, *111*, 93–107.

Bacolod, Marigee, Bernardo S. Blum, Marcos A. Rangel, and William Strange, "Learners in cities: Agglomeration and the spatial division of cognition," *Regional Science and Urban Economics*, 2023, *98*, 103838.

Charlot, Sylvie and Gilles Duranton, "Communication externalities in cities," *Journal of Urban Economics*, 2004, *56* (3), 581–613.

Ciccone, Antonio and Robert E Hall, "Productivity and the Density of Economic Activity," *American Economic Review*, 1996, *86* (1), 54–70.

De la Roca, Jorge and Diego Puga, "Learning by working in big cities," *The Review of Economic Studies*, 2017, *84* (1), 106–142.

_, Gianmarco IP Ottaviano, and Diego Puga, "City of dreams," *Journal of the European Economic Association*, forthcoming.

Glaeser, Edward L and David C Maré, "Cities and skills," *Journal of Labor Economics*, 2001, *19* (2), 316–342.

_, Jed Kolko, and Albert Saiz, "Consumer city," *Journal of Economic Geography*, 2001, *1* (1), 27–50.

Hayakawa, Kazunobu, Hans Koster, Takatoshi Tabuchi, and Jacques-François Thisse, "High-speed rail and the spatial distribution of economic activity: Evidence from Japan's Shinkansen," *mimeo*, 2021.

Higashi, Yudai, "Toshika ga roudousha ni ataeru eikyou: Roudou shijou ni okeru shuseki no keizai to fukeizai (On the impact of urbanization on workers: Agglomeration economies and diseconomies in a labor market)," *Nihon roudou kenkyu zassi*, 2022, *746.* (In Japanese).

Leonardi, Marco and Enrico Moretti, "The agglomeration of urban amenities: Evidence from Milan restaurants," *American Economic Review: Insights*, 2023, *5* (2), 141–157.

Morikawa, Masayuki, "Economies of density and productivity in service industries: An analysis of personal service industries based on establishment-level data," *The Review of Economics and Statistics*, 2011, *93* (1), 179–192.

Rosenthal, Stuart S and William C Strange, "Agglomeration and hours worked," *The Review of Economics and Statistics*, 2008, *90* (1), 105–118.

Sandvik, Jason J, Richard E Saouma, Nathan T Seegert, and Christopher T Stanton, "Workplace knowledge flows," *The Quarterly Journal of Economics*, 2020, *135* (3), 1635–1680.

Santos Silva, J.M.C. and Silvana Tenreyro, "The log of gravity," *The Review of Economics and Statistics*, 2006, *88* (4), 641–658.

Tabuchi, Takatoshi and Atsushi Yoshida, "Separating urban agglomeration economies in consumption and production," *Journal of Urban Economics*, 2000, *48* (1), 70–84.