Why Do Improvements in Transportation Infrastructure Reduce the Gender Gap in South Korea?*

Eunjee Kwon†

December 24, 2021

Abstract

This study investigates whether the increases in connectivity across cities help reduce the gender gap in labor market outcomes —a question for which there is limited prior evidence. With its high level of gender disparity, South Korea provides an ideal setting to study this issue thanks to its extensive geocoded micro-panel datasets and a massive expansion of High-Speed Rail (HSR) beginning in 2004. Using an instrumental variable strategy that leverages historical railroads constructed in Korea during the Japanese colonial era, I demonstrate empirically that the gender gap in the South Korean labor market decreased with the expansion of high-speed rail (HSR). Specifically, the gender gap in employment (wages) fell by 20% (16%) in core areas (Seoul metropolitan) and 16% (0%) in non-core regions (outside of the Seoul metropolitan). I employ a spatial general equilibrium model to structurally decompose HSR's impact into labor demand and supply channels to understand the mechanisms at play. The quantitative decomposition shows that overall, HSR increased labor demand for female-intensive jobs and decreased women's labor participation costs. Finally, the empirical evidence of the structural estimation is provided that labor demand for local service industries where women are hired more intensively increased most with HSR. In addition, in non-core areas, women's participation in the labor force was encouraged by improvement in local amenities, particularly in the areas of education and childcare, which reduced women's childcare burdens.

*I am grateful to Matthew E. Kahn, Richard K. Green, and Paulina Oliva for their guidance and encouragement as generous advisors. This paper has benefited substantially from Vittorio Bassi, Jan Brueckner, Juan Pablo Chauvin, Matthew Freedman, Roger Moon, Jorge de la Roca, Jeffrey Nugent, John Strauss, Andrii Parkomenko, Elisabet Viladecans-Marsal, and Jeffery Weaver. I thank the seminar participants at the University of Southern California Brown Bag Seminar, University of Barcelona IEB, UC Riverside Brown Bag, Seoul National University, Sogang University, University of Autonoma Barcelona, Lund University, Urban Economics Association European Meeting, Urban Economics Association American Meeting, Western Economic Association International Annual Meeting, Workshop on Neighborhoods, Externalities and Inequality, Urban and Regional Economics Workshop, Urban Economics Summer School, and Southern California Graduate Conference in Applied Economics. I thank Statistics Korea Micro Data Integrate System for providing confidential data. This paper is financially supported by a research grant from Lusk Real Estate Center and Korean Studies Institute at USC. All errors are my own.

†Department of Finance, University of Cincinnati. E-mail: eunjee.kwon@uc.edu

1 Introduction

The gender gap in the labor market narrows with economic development, yet the disparities remain salient even in developed countries like the U.S., Japan, and South Korea (Blau and Kahn (2017)). The labor force participation costs for women are generally higher, mainly due to the more substantial burden of childcare (Maurer-Fazio et al. (2011)) or home production (Greenwood et al. (2005)). As a result, women tend to seek jobs with more flexible work arrangements and choose different industries and occupations than men (Goldin and Katz (2016)).

The previous literature has shown that technological shocks can affect the gender gap in labor market outcomes. For example, technological progress, such as electrification in household sectors, reduces the domestic burden on women (Greenwood et al. (2016); Vidart (2020)) and helps reduce the gender gap in the labor market by encouraging them to join the labor force. However, on the other hand, robotization unintentionally increases the gender pay gap, as jobs that are male-dominant disproportionately benefit from robots, which ultimately increases the gender disparities in the labor market (Aksoy et al. (2019)). These findings imply that any technological progresses could potentially affect men and women differently in the labor market, as men and women face different labor market constraints and choose other industries and occupations.

This study extends the literature by investigating whether improvements in transportation infrastructure could have different labor market impacts on men and women—a question for which there is limited prior evidence. Specifically, this study investigates the impact of South Korea's high-speed rail (HSR[1]) on gender-specific local labor market outcomes. HSR is an inter-city transit which is not primarily used as a commuter train in the South Korean setting: most trips are made for visiting relatives and family, for leisure and business, or for enjoying services in other cities. As investment in such infrastructure involves considerable government spending,[2] understanding its heterogeneous impacts across different demographic groups is necessary. Despite the fast-growing literature on the impact of HSR,[3] there is only limited evidence of its impact on gender disparities

[1]HSR, a relatively recent innovation in intercity transit, has become common in both developed and developing countries. Since the construction of the first HSR, Japan's Shinkansen, in 1964, HSRs have been built across the world, including in many European as well as East Asian countries. Discussion on HSR construction is ongoing in the United States.

[2]2 For example, for the HSR in California, which links Los Angeles–San Francisco–Las Vegas, the total construction cost is estimated to be $7 billion. The total cost to construct HSR in South Korea, which links Seoul to Busan and several other major cities, is $15 billion (2015).

[3]The literature finds that HSR enhances interaction among scientists living in different cities (Dong et al. (2020)) and enhances the spatial agglomeration of cities (Zheng and Kahn (2013); see Qin (2017) for Chinese cases; for German examples, see Ahlfeldt and Feddersen (2018)). HSR also affects the location decisions of firms (Charnoz et al. (2018) for French firms), and the impacts are heterogeneous across industries (Lin (2017)). Furthermore, HSR

in labor market outcomes. This study fills this gap.

HSR is likely to affect men and women differently through both the labor demand and supply channels. On the one hand, HSR can create shocks in local gender-specific labor demand. As a transportation mode mainly for carrying passengers, HSR reduces the cost of moving people, but it does not necessarily reduce the cost of moving goods. Naturally, industries that rely more on the cost of moving people than other sectors would benefit more from HSR (Lin (2017)). If, for example, women are disproportionately hired more in such sectors, then the increases in the labor demand for female would occur, somewhat unintentionally, with HSR. The disproportional increases in labor demand for women would lead to the decreases in gender gap in labor market outcomes. On the other hand, the labor supply decisions of men and women could be affected differently. As economic activities are redistributed with HSR, amenities are endogenously changed (Gorback (2020)). Improvement in local amenities, such as in childcare facilities, would liberate women from the burden of childcare and encourage them to join the labor force. Combining both demand- and supply-side impacts, HSR could generate different labor market implications for men and women.

The massive construction of HSR in South Korea, together with the country's distinctive gender inequality and a heavy geographical concentration of economic activities, makes South Korea an ideal laboratory setting to study the research question. The gender pay gap of 34% in South Korea is much higher than the OECD average of 13.1%. Only about 45% of the women in Korea are employed, compared to 80% of men. Furthermore, men and women sort into different industries especially after marriage, as women disproportionally choose to work in local service sectors, that require less work experience and provide flexible work arrangement. Finally, economic activities are concentrated in a few districts in the Seoul metropolitan area (i.e., Seoul, Incheon, and Gyeonggi provinces). Korea Train eXpress (KTX) was first introduced in 2004, connecting Seoul with areas all the way down to the South edge of the South Korean territory.

Estimating the causal relationship between improvement in HSR and local economic outcomes, such as in employment, population, and gender gap in labor market outcomes, is complicated further by endogeneity concerns. For example, districts that expect to gain the most from HSR are more likely to be favored by the South Korean government or to lobby for HSR development more aggressively, thereby leading to the non-random placement of HSR stations. To address this

expansion can affect workers' decisions on where to live and where to work. For example, in Germany,workers prefer jobs in smaller cities while residing in bigger towns (Heuermann and Schmieder (2019)).

concern, I estimate a two-way fixed effects model using district and year fixed effects to control for unobservable variables. I leverage the fixed effects estimators with the existence of old railroad stations constructed during the Japanese invasion (1894—1945) as an instrumental variable for HSR stations' location. The identifying assumption behind this instrumental variable is that using the existing stations and routes might reduce HSR construction costs (i.e., relevance condition), and after about 100 years, the old railroad stations built during the colonial era would not impact the growth of the regional economic outcome (i.e., exclusion restriction).

My empirical findings using the old stations as instruments show that HSR reduces the gender gap in labor market outcomes. HSR narrows the gender employment gap in both core (i.e., districts in the Seoul metropolitan areas) and non-core areas (i.e., districts outside of the Seoul metropolitan areas) and decreases the wage disparities between men and women in the core areas. In addition, the expansion of HSR changes the distribution of people and jobs, decentralizing both the population and employment: population and employment increase with the KTX expansion in districts outside the core areas and decrease in the core areas.

To understand the mechanisms underlying the decrease in gender gaps with the expansion of HSR, I use a class of spatial general equilibrium models with the gender-segmented labor market, following Chauvin (2017).[4] In equilibrium, the gender gap in labor market outcomes is a function of gender-segmented industries' relative productivity and women's labor participation costs. The construction of HSR changes these two aspects for each site. The equilibrium condition allows us to decompose the mechanisms behind the HSR's impact on the gender employment and wage gaps into the labor demand channels (relative productivity) and the labor supply channels (the labor force participation costs). The quantitative decomposition combining the structural estimation and the empirical findings shows that overall, HSR increases the labor demand of female-intensive sectors and decreases women's labor participation costs. Specifically, in core areas, the labor demand impacts are more significant than the labor supply impacts, whereas the labor supply impacts are more salient in non-core areas.

Next, I empirically test the model predictions to understand the mechanisms underlying the empirical findings. First, to test relative productivity increases in female-intensive sectors, I examined how industries with different gender ratios have been affected. It is shown that male-intensive sectors, such as transportation, public administration, and manufacturing, are negatively affected

[4] "Households" in the model consist of a woman and a man who jointly decide where to live. Each member decides whether to work based on their labor force participation costs, which are higher for women. Firms hire women and men in different intermediate sectors, which are used to produce final goods.

by HSR. In contrast, female-intensive sectors such as retail, education, medical services, and restaurants are positively affected by HSR expansions. These empirical results are consistent with previous findings that HSR has different impacts across sectors. I then test whether the amenity level[5] in non-core areas increases with the HSR expansion. Focusing on the retail, education, restaurant, and medical service sectors, I show that HSR increases the amenity level in non-core areas but not in core areas. Additionally, HSR's impact on the number of workers per child in the education sector shows that education amenities in non-core areas increase significantly with HSR. This result implies that HSR potentially contributes to decreased female labor force participation costs in non-core areas by reducing the childcare burden. All these findings are consistent with the model predictions.

The contribution of this study is twofold. First, this study is related to the literature on how structural changes could have different labor market impacts on men and women. A growing body of research has investigated the effect of technological shocks on gender gaps in labor market outcomes either from the supply side (Greenwood et al. (2016); Vidart (2020)) or the demand side (Aksoy et al., 2019). This study extends the literature by examining how an improvement in transportation technology could impact both the demand and supply side of the labor market. The demand channels this study finds are similar to those in Aksoy et al. (2019): HSR reduces both the gender pay and employment gap as it increases the labor demand in female-intensive sectors (e.g., local services) more than in industries that hire men more intensively. This is also similar to how the famous Bartik shock (Bartik (1991)) works. On the other hand, women's labor supply is promoted by improvements in local amenities, particularly in education and childcare, which reduce women's childcare burden. Furthermore, this study contributes to the literature by focusing on improving transportation infrastructure as a technology shock, which few previous works document.

Second, the study is closely related to the literature on the impact of inter-city transportation infrastructure.[6] Its heterogeneous effects across spaces, [7] as well as across different socio-economic

[5]Following Diamond (2016), I define the local service amenity as the number of establishments per resident.

[6]Improvement in transit systems increases the trade of goods across spaces and reduces inter-regional price gaps (Duranton et al. (2014); Donaldson (2018); Donaldson and Hornbeck (2016)). Another strand of literature finds that transportation reduces the cost of moving people and promotes inter-regional migration (Morten and Oliveira (2016)) or changes workers' decisions on where to live and work (Baum-Snow (2007); Tsivanidis (2018); Severen (2018)). Redding and Turner (2014) provides the theoretical and empirical literature on the relationship between the spatial distribution of economic activity and transportation costs.

[7]The literature examines the differences both across and within the regions. For example, within the city-level, Baum-Snow (2007) showed that when highways in the United States connect city centers and suburbs, it decentralizes economic activities. Investigating the transportation infrastructure across cities, Baum-Snow et al. (2020) find that regional highways in China increase economic output and population in regional primates at the expense of hinterland prefectures in China.

groups (Tsivanidis (2018)), have been a focal point in recent research. This paper combines the rich transportation literature with a growing body of research in urban economics focusing on the gender differences in location choices and commuting decisions (Farré et al. (2020), Chauvin (2017), Rosenthal and Strange (2012), Black et al. (2014), Kawabata and Abe (2018), Liu and Su (2020), and Le Barbanchon et al. (2020)). While most of the discussion has focused on the contribution of commuting costs on the gender gaps in the labor markets, this paper extends the literature by focusing on non-commuting channels, which is a crucial yet under-researched aspect.

The rest of this paper is organized as follows. Section 2 provides background information on the HSR expansion in South Korea and its economic geography. Section 3 outlines the model that guides the empirical predictions. Section 4 provides a road map which links the model to empirical exercises. The data sets used and empirical methodologies in this study are described in Sections 5 and 6, respectively. Section 7 presents the empirical results of the impact of HSR on the gender wage and employment gap. Section 8 empirically tests some of the driving mechanisms derived from the model, and Section 9 quantitatively estimates the mechanisms. Section 10 presents robustness checks, and finally, Section 11 concludes.

2 Background

2.1 Economic Geography of South Korea

Home to 51 million people, South Korea is known for not only its fast economic growth but also its high urbanization rate and heavy geographical concentration of economic activities. The urbanization rate was 81.7% in 2018, which is similar to that of developed countries such as the United States, Canada, and France. Half of the South Korean population lives in the Seoul metropolitan areas[8][9]. The population of Seoul proper in 2016 was estimated at 10.29 million; however, the sprawling metropolitan area is much larger at 25.6 million. In contrast, Busan, the second-largest city in South Korea, is much smaller, with a population of 3.4 million. Figure 1 shows the spatial distribution of population and employment in South Korea.

Figure 1 about here

[8] Seoul metropolitan areas are defined as Seoul and the surrounding areas including Gyeonggi and Incheon provinces.

[9] With a population density of 16,425 per square kilometer (Statistics Korea, 2015), Seoul (the capital of South Korea) ranks second in population density among urban areas of more than 5 million people. Considering the average population density of South Korea at 528 persons per square kilometer, Seoul metropolitan areas are disproportionately large.

2.2 Gender Gap in Labor Market

The demographic structure and labor market situation in South Korea are remarkable. The gender pay gap in South Korea is the highest among OECD countries. Female full-time employees on average earn 37.2% less than a male employee as shown in Figure 2. In the labor market, gender disparities still exist even after adjusting for human capital differences. This means that the gender pay gap in South Korea likely reflects a lack of opportunity, rather than a lack of ability, for women. In addition, the disparity may be a result of other factors such as more substantial domestic burdens, discrimination, or social norms. Many women stop working when they have children. In the 2000s, fewer young people got married, and this demographic had fewer babies. The fertility rate has become close to 1 since the early 2000s, and the marriage rate has halved compared to that of the 1980s.

Figure 2 about here

Figure 3 presents the gender differences in time use over the years (Korea Income and Labor Panel Survey, 2000—2015). Female employment has seen slight increases, but the gender gap has not narrowed significantly. Around 80% of men responded that "working full-time" is their main activity, whereas only 40—50% of women answered so, despite the slight increases over the years. Another 40—50% of women answered "doing domestic work" as their main activity, and about 10% of women partly worked and did domestic work, which has decreased over time. Domestic burdens do not seem to be a significant concern for men. Less than 5% of men answered that domestic work was their primary activity with almost no changes in this number over the years.

Figure 3 about here

In South Korea, gender wage gaps have been consistent over the years, even after adjusting for the human capital (education) and experience level. Figure 4 shows gender differences in wages over the years (Korea Income and Labor Panel Survey, 2000—2015). The unadjusted wage differences between men and women have remained at around 40—45% over the years. Adjusted wage differences have become even more considerable and persistent at around 40—44%.

Figure 4 about here

Figure 5 shows the spatial distribution of sex ratio in employment, defined as male employment over female employment. Overall, the core areas or big cities have a higher male-to-female employment ratio in 2000. Changes in the sex ratio in Panel B show that female employment increased more than male employment, especially in core areas.

Figure 5 about here

Table 1 presents distinctive demographic characteristics in core and non-core areas. In Panel A., differences in labor force participation rate between men and women are significant. The share of working men does not differ between core and non-core groups, and approximately 75.5% of men work. However, the labor force participation rate for women is much lower than that for men, and even lower in the core areas (39.72%) than in non-core areas (46.96%). Finally, for both women and men, a disproportionately high number of singles live in core areas, which is consistent with the marriage market hypothesis.

In terms of gender disparities in labor market outcomes (Panel B), the employment gap is slightly larger in core areas, whereas the wage gap is slightly higher in non-core areas. However, overall, the gender gap in both intensive and extensive margins is not negligible.

Table 1 about here

2.3 Why Women Sort into Different Sectors from Men?

Understanding what causes women and men to be sorted into different sectors in South Korea is a topic of interest. Figure 6 shows the life cycle of men and women in terms of their marriage and working status. Men and women tend to show similar labor force participation rates until their mid-twenties. However, in their mid-twenties, as women begin getting married, they tend to exit the labor market due to the childcare and other domestic burdens. Some women reenter the labor force in their mid-thirties, but the number of women never reaches that of men.

Figure6 about here

Moreover, women tend to work in different sectors or occupations when they reenter the labor market after marriage. Figure 7 compares occupation compositions of each age group of men and women before their exit from the market (ages 25–34) and after reentry (ages 35–44). Women and men at earlier life stages (ages 25–34) show a similar pattern in the compositions of occupations; generally, they work in white-collar jobs such as managers, professionals, or other office workers. However, after women's return to the labor market at the age of 35–45, a significant number of women work in sales and local services (Figure 7, (b) and (d)) compared to that of men, which does not change much in terms of occupation compositions. This shows that an inordinate number of women work in local service sectors after their marriage as the jobs require relatively less work experience and provide more flexibility.

Figure7 about here

This trend is related to the work of Goldin and Katz (2016), which shows that the gender gap in the pharmacy sector decreased over time in the United States, as the jobs have less penalty in the pay for part-time work and meet women's demand for flexibility in working hours.

2.4 High-Speed Rail Expansion

Korea Train eXpress (KTX), the HSR in South Korea, was first introduced in 2004 and experienced massive expansion from 2010 to 2012. The construction plans for KTX were first made in the 1980s to reduce both unequal development across spaces and the problems that became acute in areas near Seoul, such as unaffordable housing problems and traffic congestion.

The bullet train not only links Seoul and Busan but also connects small towns, which were previously only poorly connected to big cities. The KTX expansion process is depicted in Figure 8 and Table A2. In 2004 (stage 1), the Gyeongbu-line, which connects Seoul to the southeastern region, and the Honam-line, which connects Seoul to the southwestern part of South Korea, were opened. In 2010–2012 (stage 2), the Jeonla-line began operation, the Gyeongbu-line was extended from Daegu to Busan, and some stations in Gyeongjeon opened. In 2012, KTX had covered 22% of the total territory and served 56% of the total South Korean population (KOSIS).

Figure 8 about here

Table A2 about here

KTX makes round-trips between cities feasible for single-day travel. Travel from Seoul to Busan (417.4 km), two metropolitan areas located at opposite edges of South Korea, takes 6 hours by car or non-high-speed train, but only around 2 hours by KTX at the fastest speed. The train is mainly used for transporting people rather than goods. More than 80% of KTX passengers use KTX to visit friends or relatives or to go on business trips. Fewer than 1% of the passengers use KTX for their daily commutes because of the expensive ticket price.[10]

[10] A one-way train ticket from Seoul to Busan is around $40. As the average household net-adjusted disposable

As observed in Figure 9 (a), KTX ridership has increased significantly since its opening. KTX accounted for 3.7% of interregional ridership (buses, cars, KTX or non-KTX train) in 2004 and reached 10% in 2015[11], while ridership on other transportation modes (except for domestic flights) has decreased over time. Considering the total passenger-km riderships in Figure 9 (b), the increase for KTX is even more significant. Total passenger-km of KTX outweighed that of non-KTX trains from 2006, which drives the increases in total passenger-km from 2011. Combining 9 (a) and (b), in 2015, average travel distance per ticket = 15,000,000,000 (passenger-km) / 60,000,000 (n. tickets)= 250 km, which is certainly over the average commute distance in South Korea.

Figure 9 about here

The direct competitors of KTX are intercity buses or existing train networks. However, in South Korea, domestic flights are not considered significant interregional transits as airports are located far from city centers. Therefore, flying is regarded as a less-efficient means of transit.[12]. Indeed, the ridership on any domestic flights (disregarding flights to Jeju Island) was almost stagnant during the sample period of 2000–2015, whereas ridership for non-KTX trains or intercity buses decreased significantly over the sample period.

3 Model

This section provides a model for explaining how the expansion of HSR can affect the gender gap in the labor market and the distribution of economic activities. Here, the standard spatial general equilibrium framework as Roback (1982) is utilized; with local wages, housing rents, and amenities. The general equilibrium approach captures both the direct effects of the shock as well as endogenous adjustments of factor prices and quantities (Moretti (2010)). Besides, I borrow the notion of the gender-segmented labor market and location choice of a married couple framework from Chauvin (2017). I advance the idea by allowing HSR to change the equilibrium outcomes through the changes in (1) productivity of different intermediate goods, (2) amenities of the locations, and (3) labor force participation costs of females.

---

income per capita was USD 19,372 per year in 2015 (USD 1,614.33 per month), train tickets are too expensive to be used for daily commutes.

[11]Korea Transport Database of the Korea Transport Institute (KTDB).

[12]Speedy trains and planes are generally competitive if travel time is less than 1,000 kilometers (621 miles). (Bloomberg, 2018)

In the model, a household, consisting of a wife and a husband, chooses where to live (and they work in the same location). Each household member decides whether to work or not, given their labor force participation costs and the wage level at each location. As is well known in the literature, the cost of labor force participation is higher for women.[13] Women and men are hired by different local industries that produce intermediate goods. These intermediate goods are aggregated locally with imperfect substitutions as a nationally produced good. Finally, housing is supplied locally. A spatial equilibrium consists of a set of local wage levels for each gender, housing prices, a mapping of households to cities, locations, and sectors so that firms maximize profits, households maximize utility, and labor markets in each gender-segmented sector and housing markets are clear.

3.1 Location Choice of Households

Indirect utility of a household i, living in j at t[14] is as follows:

$$U_{jt} = \alpha + w_{jt}^{net} - r_{jt} + A_{jt} \tag{1}$$

with total household earning ($w_{jt}^{net} = w_{jt}^{M,net} + w_{jt}^{W,net}$, where $w_{jt}^{G,net}$ indicates local wage level of each gender group $G \in \{Women(W), Men(M)\}$), rent ($r_{jt}$), amenities ($A_{jt}$), and some constant α. Specifically, local wage of each gender group ($w_{jt}^{G,net}$) is as follows:

$$w_{jt}^{G,net} = \begin{cases} w_{jt}^{G} - \phi^{G} & if w_{jt}^{G} > \phi^{G} (working) \\ 0 & if w_{jt}^{G} < \phi^{G} (notworking) \end{cases} \tag{2}$$

with the cost of labor force participation, $\phi^W > \phi^M$, which makes men more likely to work.

Following Chauvin (2019), men and women draw an exogenous and stochastic labor force participation cost, after they move to a new region. The CDF of labor participation costs follow a power law ($F(\phi_i) = (\frac{\phi_i}{\phi_{min}})^{\iota}$). Women's labor participation cost CDF has higher support: the support for men is $\phi^M \in (1, \phi_{max})$ and for women is $\phi^W \in (1 + T_{jt}, \phi_{max})$, for $T_{jt} > 0$. The supports for men and women are determined locally. (Wo)men will work if $\phi_i < W_{Gjt}$. This gives female labor supply function of $N_{Wjt} = N_{jt}(\frac{W_{Wjt}}{1+T})^{\iota}$, and male labor supply function of $N_{Mjt} = N_{jt}(W_{Mjt})^{\iota}$.

[13] The higher cost of labor force participation for women could be due to commuting, childcare, or home production.

[14] The indirect utility function is driven from a household's Cobb-Douglas utility function. Household collaboratively derive utility from the consumption of tradable good (C_{jt}) priced one, housing (H_{jt}) rented at r_{jt}, and local amenities (A_{jt}).

Wage gap from female and male labor supply functions is:

$$\frac{W_{jt}^M}{W_{jt}^W} = \frac{1}{(1+T_{jt})}(\frac{N_{Mjt}}{N_{Wjt}})^{\frac{1}{\iota}} \tag{3}$$

3.2 Firm: Gender-Segmented Intermediate Good Industries

In the gender-segmented labor market, women and men (G $\in \{Women(W), Men(M)\}$) are hired in different intermediate good sectors:

$$Y_{Gjt} = \Psi_{Gjt} N_{Gjt}^{\beta} K_{jt}^{1-\beta}$$

Intermediate goods are combined with constant elasticity of substitution into a nationally-traded final good priced:

$$\begin{aligned} Y_{jt} =& [(Y_{Mjt})^{\sigma} + (Y_{Wjt})^{\sigma}]^{\frac{1}{\sigma}} \\ \equiv Y_{jt} =& [(\Psi_{Mjt} N_{Mjt}^{\beta})^{\sigma} + (\Psi_{Wjt} N_{Wjt}^{\beta})^{\sigma}]^{\frac{1}{\sigma}} K_{jt}^{1-\beta} \end{aligned}$$

This gives the following labor demand equations in each sector:

$$\begin{aligned} W_{jt}^G &= \beta \Psi_{Gjt} N_{Gjt}^{\beta\sigma-1} L_{jt}^{\frac{1-\sigma}{\sigma}} \\ where L_{jt} &= [(\Psi_{Mjt} N_{Mjt}^{\beta})^{\sigma} + (\Psi_{Wjt} N_{Wjt}^{\beta})^{\sigma}]^{\frac{1}{\sigma}} \end{aligned}$$

and the local gender wage gap is:

$$\frac{W_{jt}^M}{W_{jt}^W} = (\frac{\Psi_{Mjt}}{\Psi_{Wjt}})^{\sigma} (\frac{N_{Mjt}}{N_{Wjt}})^{\beta\sigma-1} \tag{4}$$

3.3 Housing Market

The housing supply closely follows that in Glaeser et al. (2008). In every location j, there is an absentee landlord who buy the housing from developers and rent it to local residents at r_{jt}. Profits for the developers are given by:

$$\pi_{jt} = \Sigma_{t=0}^{\infty} \frac{r_{jt}}{(1+i_t)} - CC_{jt} \tag{5}$$

where i_t is the national interest rate and CC_{jt} are the local construction costs. The free entry and zero-profit conditions give the local housing supply.

The housing demand function is from the household's Cobb-Douglas utility function, which is aggregated at the number of households located in location j at time t (N_{jt}). In equilibrium, the local housing market clears for every location j at time t.

3.4 Equilibrium: Gender Gap in Labor Market

Equating labor supply (equation 3) and labor demand (equation 4) functions, in equilibrium, the following equations regarding the gender gap in labor market outcome hold.

The gender employment gap (male-to-female ratio, $\frac{N_{jt}^M}{N_{jt}^W}$) becomes

$$\frac{N_{jt}^M}{N_{jt}^W} = (1 + T_{jt})^{\Omega} (\frac{\Psi_{Mjt}}{\Psi_{Wjt}})^{\sigma\Omega} \tag{6}$$

The gender wage gap (male-to-female ratio, $\frac{W_{jt}^M}{W_{jt}^W}$) becomes

$$\frac{W_{jt}^M}{W_{jt}^W} = (1 + T_{jt})^{(\beta\sigma-1)\Omega} (\frac{\Psi_{Mjt}}{\Psi_{Wjt}})^{\frac{\sigma}{\iota}\Omega} \tag{7}$$

where $1 + T_{jt}$ is labor force participation cost for women in location j at time t, $\frac{\Psi_{Mjt}}{\Psi_{Wjt}}$ is relative productivity between male- (Ψ_{Mjt}) and female-intensive sector (Ψ_{Wjt}). Given the range of parameters; the labor share in Cobb-Douglas production of the intermediate good (β), the elasticity of substitution of final good production (σ), and the CDF parameter of the female labor force participation cost function (ι), $\Omega = \frac{1}{\frac{1}{\iota} - \beta\sigma + 1} > 0$ and $(\beta\sigma - 1)\Omega < 0$.

The two equilibrium equations in the gender gap provide intuitive interpretations. Note that in Equation 6 and 7, parameters in blue is positive value, and in red is negative value. In Equation 6, the gender employment gap increases with labor force participation cost for women. As women participate in the labor force only if the market wage is higher than their costs, higher labor force participation cost lowers women's labor participation rate, which leads to the increases in the gender employment gap. On the other hand, as in Equation 7, the higher cost would increase the overall wage level of women who join the labor force, which decreases the gender wage gaps. Meanwhile, the increase in relative productivity between male- and female-intensive sector would worsen the gender gap in both employment and wage, as the labor demand for men increases more than that for women. Equations 6 and 7 give us the following predictions regarding the gender employment and employment gap.[15]

3.5 Prediction from the Model

Table 2 summarizes the changes in the gender employment and wage gap, when relative productivity and female labor participation cost change simultaneously due to HSR expansion. For example, if both the relative productivity of the male-intensive sector and the female labor force participation cost decrease, the gender gap in employment would decrease. Simultaneously, the changes in the gender wage gap would be indeterminate.

Table 2 about here

3.6 Comparative Statics: Impact of HSR

With HSR, location-specific amenities, productivities, and the labor force participation cost for women could be affected:

1. Changes local amenities is expressed as follows:

$$U_{ijt} = \alpha + w_{jt}^{net} - r_{jt} + \theta_{jt}$$

2. Changes in relative productivity of intermediate goods is expressed as follows:

$$\frac{W_{Mjt}}{W_{Wjt}} = (\frac{\Psi_{Mjt}}{\Psi_{Wjt}})^{\sigma}(\frac{N_{Mjt}}{N_{Wjt}})^{\beta\sigma-1}$$

3. Changes in labor force participation cost for women is expressed as follows:

$$\frac{W_{jt}^{M}}{W_{jt}^{W}} = \frac{1}{(1+T_{jt})}(\frac{N_{Mjt}}{N_{Wjt}})^{\frac{1}{\iota}}$$

4 Road Map: From Model to Empirics

Taking derivatives of Equation 6 and 7 with respect to HSR, the following equations are obtained.

For the changes in the gender employment gap:

[15]Population in area j at time t is N_{jt} can be expressed with respect to all the exogenous parameters and local amenities (θ_{jt}), relative productivity of intermediate good sectors ($\frac{\Psi_{Mjt}}{\Psi_{Wjt}}$), and labor participation costs of women (T_{jt}).

$$\frac{\partial(\frac{N^M_{jt}}{N^W_{jt}})}{\partial HSR} = \Omega\frac{\partial(1+T_{jt})}{\partial HSR} + \sigma\Omega\frac{\partial(\frac{\Psi_{Mjt}}{\Psi_{Wjt}})}{\partial HSR} \tag{8}$$

and for the gender wage gap:

$$\frac{\partial(\frac{W^M_{jt}}{W^W_{jt}})}{\partial HSR} = (\beta\sigma - 1)\Omega\frac{\partial(1+T_{jt})}{\partial HSR} + \frac{\sigma}{\iota}\Omega\frac{\partial(\frac{\Psi_{Mjt}}{\Psi_{Wjt}})}{\partial HSR} \tag{9}$$

where $\Omega = \frac{1}{\frac{1}{\iota} - \beta\sigma + 1}$ and equation in blue is observable, and in orange is unobservable.

Empirically, changes in local employment gap ($\frac{\partial(\frac{N^M_{jt}}{N^W_{jt}})}{\partial HSR}$) or local wage gap ($\frac{\partial(\frac{W^M_{jt}}{W^W_{jt}})}{\partial HSR}$) in Equation 8 and 9 can be directly observed. However, the relative productivity between male- and female-intensive sectors ($\frac{\partial(\frac{\Psi_{Mjt}}{\Psi_{Wjt}})}{\partial HSR}$) or labor force participation cost for women ($\frac{\partial(1+T_{jt})}{\partial HSR}$) are not easily observed from the data. Additionally, the range of the three parameters (β, σ, and ι) are determined in the model.

Given the availability of data and parameters, the model predictions to the empirics were linked as follows:

1. Section 7 empirically estimates the impact of HSR on the local gender gaps in the labor market outcomes ($\frac{\partial(\frac{N^M_{jt}}{N^W_{jt}})}{\partial HSR}$, $\frac{\partial(\frac{W^M_{jt}}{W^W_{jt}})}{\partial HSR}$), which are directly observed from the data.

2. Section 8 infers the impact of HSR on the relative productivity and female labor force participation costs based on the results of Section 7 and the model predictions in Table 2.

3. Section 8 continues to investigate whether the inferred predictions can be confirmed with suggestive evidence from the data.

4. Finally, Section 9 structurally decomposes the impact of HSR on the relative productivity ($\frac{\partial(\frac{\Psi_{Mjt}}{\Psi_{Wjt}})}{\partial HSR}$) and labor force participation cost for women ($\frac{\partial(1+T_{jt})}{\partial HSR}$), based on the calibrations and the results in Section 7.

5 Data

This section describes the data sets considered in analyzing the causal impacts of the massive bullet train expansion on South Korea's economic geography. To construct a geocoded 16-year panel data set at the district-level, the Census on Establishment, the Population and Housing Census, and the Internal Migration Statistics collected by Statistics Korea are utilized. These administrative data sets are combined into a time-varying travel-time matrix, which is constructed using the algorithm explained in the Appendix. Note that the microlevel data for the Census on Establishment and the Population and Housing Census are exclusively provided through the subscription to the Micro Data Integrated System (MDIS), operated by Statistics Korea. Throughout this study , regions are defined at district-level. In 2010, there were 228 districts consisting of 15 provinces[16]. By combining the data sets explained above, a (nearly) balanced panel data set at the district-level (with the number of districts equal to 228) from 2000 to 2015 (t=16) was constructed. The details of each data set are described in Appendix A. Summary statistics are presented in Table 3.

Table 3 about here

6 Empirical Strategy

The following hypotheses from the model are empirically tested in this section. Does the construction of HSR affect the gender gaps in the labor market outcomes, such as wage levels and employment? Does the construction of HSR shift population and employment from core to non-core areas or does it induce more concentration of economic activities in core areas?

6.1 OLS Regression

In a district i within a province s at time t, the econometric model estimating the impact of the KTX expansion on the outcome variable of interest is:

$$log(y_{it}) = \alpha + \beta Treat_i Post_t + \delta X_{it} + \eta_{st} + f_i + \phi_t + \epsilon_{it}, \tag{10}$$

where y_{it} is a dependent variable including the gender gap in labor market outcomes, employment, population, or wage-level of a county i of a province s at time t. $Treat_i Post_t$ is the treatment dummy equal to one, if the centroid of a district i is within 15km of any KTX station at time t. A set of control variables (X_{it}), *district fixed effect* (f_i), *year fixed effect* (ϕ_t) and *province-time varying fixed effect* (η_{st}) are included. β can be interpreted as the percentage change in the outcome of interest with new construction of a KTX station in a district. Here, all the standard errors are

[16] As our main interests are railroad transportation, districts on Jeju Island and Ulleng Island were excluded in this paper.

clustered at district-level. Instead of using the spatial center as centroids, this study calculates the population-weighted centroids of each districts, based on the zip-code level population data. Refer to Appendix for the graphical representation of the calculation.

To identify the heterogeneous impacts between core and non-core areas, consider the following specifications:

$$log(y_{it}) = \alpha + \beta_c Treat_i Post_t D_i^c + \beta_{nc} Treat_i Post_t D_i^{nc} + \delta X_{it} + \eta_{st} + f_i + \phi_t + \epsilon_{it}, \quad (11)$$

where $D_i^c(D_i^{nc}) = 1$ if a district i is located in the core (non-core). In terms of the institutional setting in South Korea, core areas are defined as districts located in the Seoul metropolitan areas (Seoul, Gyeonggi, and Incheon provinces) and non-core areas are districts in the rest of the provinces.

6.2 Instrumental Variable

6.2.1 Non-random Placement of HSR

Studying the causal effect of transportation requires careful inspection due to selection issues as this is the nature of place-based policies. The central inference problem researchers are facing in the transportation literature is that transportation is not assigned randomly but is determined based on both observed or unobserved location characteristics and forecasts of expected benefits (Redding and Turner (2014)).

In the context of this study, although the location of high-speed train stations in South Korea was determined by the central government 14 years prior to the actual opening (Baek and Park (2015)), possible endogeneity concerns remain. For example, districts that expect to gain the most from HSR are more likely to lobby more aggressively to have the station in their districts . This makes the location choice of HSR nonrandom. The unobserved factors that could affect the actual timing of the opening can compromise the identification of the regression. To address these issues, this study suggests the following instrumental variable estimation, following the literature in urban economics.

6.2.2 Instrumental Variable: Old Railroad Stations in Japanese Colonial Rule

Existing literature has dealt with this endogeneity issue by using either planned (Baum-Snow (2007); Duranton and Turner, 2012; Duranton et al. (2014)) or historical (Morten and Oliveira

(2016); Garcia-López et al. (2015); Tsivanidis (2018)) railroads, or least-cost corridors (Faber (2014)) as instrumental variables. In this study, the historical railroad stations, constructed during Japanese colonial rule (before 1945) were used as the instrumental variable for the HSR stations.

In 1894, Japan assumed railroad construction in the Korean territory as part of their "modernization process." The first railroad network in Korea, Gyeongin-line, was opened in 1899. From then, the Japanese railway ministry expanded railroad networks over the years. From 1904–1906, the Gyeongbu-line and the Gyeongui-line were added to serve as military pathways to China and Russia; in 1910, the Honam-lines were added to help transport natural resources and agricultural products from South Korea to Japan.

Figure 10 about here

The assumption behind using the historical railroad stations as instrumental variables for HSR is that the existing station platforms and routes might be more cost efficient, which satisfies the relevance condition of IV($Pr(D=1|Z=0) \neq Pr(D=1|Z=1)$), but at the same time they might have no impact on current economic situations in the counties (exclusion restriction, $E(\epsilon|Z=1) = E(\epsilon|Z=0)$).

Table 4 shows that the relevance condition of the IV is satisfied. The first stage of the regressions in Table 4 shows high F-statistics, which are above the typical rule of thumb (F-statistic of more than 10, Stock and Yogo (2002)). The high F-statistics confirm that the existence of the old railroad stations in districts predicts a higher probability of the construction of HSR in the districts.

Table 4 about here

On the exclusion restriction, old railroad stations, constructed before 1945, in the Japanese colonial period, would affect the growth of the economies of counties in the early 2000s. Then, trains were mainly used to transport resources from South Korea to Japan or as assets of the Russian-Japanese war, rather than for supporting the economic development of the region. Even if one argues that the existence of old railroad stations initially determines the economic situation of counties, the district fixed effects (f_i) control for such an initial level difference across counties. Here, the identification strategy comes from the condition that the old railroad station might have an impact on the level of the outcome variables, but not on their growth.

6.2.3 IV estimation

As the instrument variable is time-invariant (Z_j), the long-difference between the final year (2015) and the initial year (2000) of Equation 10 is computed as follows:

$$\Delta log(y_i) = \beta \Delta Treat_i + \delta \Delta X_i + \Delta \eta_s + \Delta \epsilon_j \tag{12}$$

Here, $Treat_i$ with $Z_i = 1$ is used as the instrument variable, if a railroad station was built in district i during the Japanese colonial era.

Finally, for capturing heterogeneous impacts across locations, as in Equation 11, $treat_i post_t D_j^k$ is instrumented by $Z_i D_i^k$, for $k \in \{core, non-core\}$.

7 Main Results

7.1 Effect of HSR construction on Local Employment and Population

The OLS regression results of Equations 10 and 11 are presented in Table 5. If all the coefficients are correctly identified, each coefficient can be interpreted as the percentage change in the outcome variables when the HSR station is located within 15 km of the centroid of districts. In columns (1)—(3) of the table, any aggregated treatment effects of HSR on population, employment, or the number of establishments cannot be observed. Conversely, once the heterogeneous treatment effects across core and non-core districts are considered (columns (4)–(6)), we observe that HSR construction is associated with 7.3% and 3.9% increases in population and employment in non-core areas, respectively, and 5.6% (with no statistical significance) and 13.6% decreases in population and employment in core areas, respectively. The results show that HSR induces the redistribution across space by moving economic activities from the core areas to non-core areas.

The OLS regression results of equation 10 and 11 are presented in table 5. If all the coefficients are correctly identified, each coefficient can be interpreted as the percentage change in the outcome variables when the HSR station is located within 15km of the centroid of districts. In column (1)-(3) of Table 5, we do not see any aggregated treatment effects of HSR on population, employment, or the number of establishments. In contrast, once we consider the heterogeneous treatment effects across core and non-core districts (column (4)-(6)), we see that HSR construction is associated with 7.3% and 3.9% increases in population and employment in non-core areas respectively, and 5.6% (with no statistical significance) and 13.6% decreases in population and employment in core areas respectively. The results show that HSR induces the redistribution across space, by moving economic activities from the core areas to the non-core areas.

Table 5 about here

Table 6 summarizes the IV estimations using the old railroad stations as instrumental variables. The qualitative results of IV estimates are consistent with the OLS estimates in Table 5, with magnitudes of two to three times and more statistical strength. Meanwhile, areas with high economic concentrations lose population, employment, and numbers of establishments, whereas the less-developed provinces benefit from increased population and jobs after HSR stations open.

Table 6 about here

We can infer the direction of the bias by comparing the OLS and the IV estimates. Upward and downward biases in OLS estimates are shown in core and non-core areas, respectively. These biases mean that HSR stations were constructed in the districts that are expected to grow in core areas, whereas in non-core areas, the districts that are expected to shrink were likely to get HSR stations.

7.2 Impact of HSR on Employment and Wages by Genders

Table 7 shows the heterogeneous treatment effects across genders in population and employment. The female and male population are affected in similar ways both quantitatively and qualitatively. However, in terms of the sex ratio in employment (i.e., male employment over female employment), female employment is observed to greatly increase in both core and non-core areas with HSR expansions.

Table 7 about here

The effects of HSR on land price and wages are reported in Table 8. Land prices in non-core areas increased by 5.8%, whereas those in core areas do not change with HSR expansion. Interestingly, an examination of the HSR's differential wage impacts across gender shows that the female wage level in core areas increases by approximately 20.3% in core areas, which contributes to the decrease in the wage gap in these core areas, by 15.8%. Conversely, was no changes were observed in the wage level in non-core areas, both for women and men, resulting in no changes in the wage gap in these non-core areas.

Table 8 about here

8 Mechanism

Table 9 summarizes the empirical findings of the impact of HSR are. HSR expansions in South Korea reduce the concentration in core areas while increasing the population and employment rate in non-core areas. Moreover, interesting gender differences for labor market outcomes are found. HSR contributes to the decrease in gender earning disparity in core areas. Moreover, the employment of women is higher in both core and non-core areas, compared to that of men.

Table 9 about here

8.1 Predictions from the Model

Combining the empirical findings in Table 9 and the predictions from the model presented in Table 2, the impact of HSR on female labor force participation costs and relative productivity between female- and male-intensive sector can be predicted in core and non-core areas.

1. In the core areas, in order to have the decrease in the male-to-female wage gap, but no changes in the gender gap in employment, the relative productivity of female intensive sectors should increase ($\frac{\partial \frac{\Psi_{Mjt}}{\Psi_{Wjt}}}{\partial HSR} < 0$). The impact on the labor cost participation cost is indeterminate, which is an empirically testable question, ($\frac{\partial T_{jt}}{\partial HSR}?0$). See table A3 for proof.

2. In the non-core areas, in order to have no changes in the male to female wage gap, but the increase in the employment rate of female, the relative productivity of female intensive sectors should increase ($\frac{\partial \frac{\Psi_{Mjt}}{\Psi_{Wjt}}}{\partial HSR} < 0$) and impact on the labor cost participation cost should decrease with the expansion of HSR ($\frac{\partial T_{jt}}{\partial HSR} > 0$). See table A4 for proof.

The following sets of regressions empirically test the propositions.

8.2 Impact of HSR Across Industries with Different Gender Intensity

Whether HSR increases the relative productivity of female-intensive sectors compared to male-intensive sectors was verified. Table 10 presents the effect of HSR across different sectors with different gender intensities in employment (i.e., male-to-female employment ratio). The most male-intensive industry is transportation, with a ten-to-one male-to-female gender ratio, whereas the most female-intensive sectors are the lodging and restaurant industries, with a male-to-female ratio of 0.5. Interestingly, sectors with higher male workers intensity, in general, show no effects on the growth in their employment, except for the construction industry. On the other hand, the increase in employment in non-core areas was shown in most of the female-intensive sectors such as retail, education, medical, and restaurant industries.

Table 10 about here

Table A5 investigates whether the changes in sex ratio within industries with HSR expansion are not driven by the changes in sex ratio. Sectors positively affected with HSR, such as retail, education, medical service, and restaurant industries, do not show extreme changes in male-to-female employment ratio with the HSR expansion, except for the education sector in non-core areas. In summary, this confirms that HSR positively affects female-intensive sectors compared to the male-intensive sectors, thus increasing labor demand for women.

Table A5 about here

8.3 Impact of HSR on Endogenous Amenity

Whether HRS has effects on the local amenity level was verified. Specifically, following Diamond (2016), local amenity level is defined as the number of establishments per resident of certain service industries. Table 11 shows the impacts of HSR on the local amenity level. Here, the total establishments of retail, medical services, and education sectors per resident in non-core areas are positively affected in non-core areas, with HSR expansion. Conversely, the local amenities are not affected by HSR in the core areas.

Table 11 about here

Notably, the education amenity level increases significantly in non-core areas. Specifically, whether workers working in the education sector per kids (under 13 years old, which is elementary school age) changes with HSR has been verified in column (5) of Table 11. Interestingly, an increase in workers per kid in the education sector is only observed in non-core areas. This indirectly implies that childcare burden for women decreases with HSR, which can potentially reduce the labor force participation costs of women.

8.4 Impact of HSR on Migration Decisions of Singles and Couple

Finally, whether the HSR changes the demographic compositions (single vs. couples) in different areas was verified. The composition changes in demographics indirectly indicate location characteristics, with the revealed preference argument. For example, if singles and couples are segregated in different regions, this indicates that each region provides a location benefit favored by different demographic groups.

The migration flow of individuals and households with different marital statuses was examined. Population is a stock variable that impairs our understanding of the underlying migration flows of individuals. Hence, the impact of the reduction in travel time between origin (o) and destination (d) counties due to KTX networks on the number of migrants for each county-pair, which provides a better idea on the movement of people. As reduced travel time improves both physical and emotional accessibility between districts, individuals more easily sort themselves into more favorable locations.

8.4.1 Data

The data set mainly used in this section is from the Internal Migration Statistics. From the household level of 88,248,353 observations over 15 years (2001—2015), origin–destination county-pair level data is aggregated, which consists of 849,600 origin–destination pairs. Additionally, to observe the various patterns across the population, this study investigates whether the migration patterns are affected differently by travel time among different marital status groups (singles vs. couples). The details of the data set are provided in Appendix A.

8.4.2 Empirical Strategy: Gravity Model of Migration Flow

A gravity model guides us in analyzing the relationship between changes in travel time and number of migrants. The model is often used to explain the commuting patterns (Heuermann and Schmieder (2019)) as well as the migration patterns (Morten and Oliveira (2016)) between regions. From the model, each empirical specification is derived as follows. The details of the model derivation are provided in Appendix C

Considering the potential endogeneity issue, the main regression equation is expressed as[17]

$$log(M_{odt}) = \alpha + \gamma log(TravelTime_{odt}) + \eta_{od} + \nu_{ot} + \theta_{dt} + \epsilon_{odt} \tag{13}$$

where, $log(TravelTime_{odt})$ is the travel time between origin and destination counties at time t, η_{od} is the origin-destination pair fixed effect, η_{ot} is the origin-year fixed effect, and θ_{dt} is the

[17] The same empirical specification is also used in Morten and Oliveira (2016).

destination-year fixed effect.

Fixed effects of the regression control for the potential endogeneity in the placement of HSR, which are parsimonious enough to capture those unobserved heterogeneities. η_{od} captures the systematic time-unvarying relationship between origin and the destination counties; time-varying origin (η_{ot}) and destination (η_{dt}) fixed effects capture all potential unobserved changes in origin and destination counties. For example, changes in inflow or outflow of migrants, due to an economic downturn or boom of origin (or destination), can be controlled by the time-varying fixed effects. Finally, ϵ_{odt} is clustered at the origin–destination pair-level.

The specific interests in this study are whether the migration flow between core areas and non-core areas are affected once HSR reduces travel time between the two areas. To capture the heterogeneity in the elasticity of migration considering the travel time reduction, the three interaction terms are placed in the regression: (origin: core; destination: non-core), (origin: non-core; destination: core), and (both origin and destination are (non)-core). Thereafter, the regression becomes

$$\begin{aligned} log(M_{odt}) =& \alpha + \gamma log(TravelTime_{odt})[\gamma_{SN} D_{o:S,d:N} + \gamma_{NS} D_{o:N,d:S} + \gamma_{within} D_{o,d:within}] \\ &+ \eta_{od} + \nu_{ot} + \theta_{dt} + \epsilon_{odt} \end{aligned} \tag{14}$$

The changes in the number of migrants in response to the changes in travel time (i.e., migration elasticity with respect to travel time changes) for each direction are follows:

- γ_{SN} : Elasticity of migration from core to non-core districts
- γ_{NS} : Elasticity of migration from non-core to core districts
- γ_{within}: Elasticity of migration within core counties or non-core districts

8.4.3 Results

Elasticities of Migration with Respect to the Travel Time Reduction

The estimation results of Equation 14, for each demographic group are presented in Tables 12 and 13.[18]

[18] In the data, we cannot determine the legal relationship between people, who move together. Men and woman who have moved in together are considered couples if both individuals are older than 18 and if their age differences are less than 12 years.

Table 12 about here

Table 13 about here

As travel time is reduced by the bullet train, single households generally migrate to the core areas (Table 12), whereas couples move to non-core areas (Table 13), leading to demographic segregation. This result might be related to the marriage market hypothesis that single households, who are active in the marriage market, are more likely to commit to or move to core areas, which have a larger marriage market pool. In contrast, couples with children moved out from core to non-core areas. One of the reasons behind the couples' movement is that couples might want to consume larger spaces as it is shown in column (3) of Table A6, so they move out to non-core areas as they are cheaper.

9 Quantifying the Mechanism: Structural Model

Let's recall Equation 8 and 9 in Section 3.

For the changes in the gender employment gap,

$$\frac{\partial(\frac{N_{jt}^{m}}{N_{jt}^{fe}})}{\partial HSR} = \Omega\frac{\partial(1+T_{jt})}{\partial HSR} + \sigma\Omega\frac{\partial(\frac{\Psi_{Mjt}}{\Psi_{Wjt}})}{\partial HSR} \tag{15}$$

and for the gender wage gap,

$$\frac{\partial(\frac{W_{jt}^{m}}{W_{jt}^{fe}})}{\partial HSR} = (\beta\sigma-1)\Omega\frac{\partial(1+T_{jt})}{\partial HSR} + \frac{\sigma}{\iota}\Omega\frac{\partial(\frac{\Psi_{Mjt}}{\Psi_{Wjt}})}{\partial HSR} \tag{16}$$

where $\Omega = \frac{1}{\frac{1}{\iota}-\beta\sigma+1}$.

The structural decomposition using Equation 15 and 16 is as follows. $\frac{\partial(\frac{W_{jt}^{m}}{W_{jt}^{fe}})}{\partial HSR}$ and $\frac{\partial(\frac{N_{jt}^{m}}{N_{jt}^{fe}})}{\partial HSR}$ can be directly estimated, from the reduced form estimations using the instrumental variables in Section 7. Under the assumptions of a set of parameters (ι, β, σ), Equation 15 and 16 become a system of two linear equations with two unknown variables, $\frac{\partial(1+T_{jt})}{\partial HSR}$ and $\frac{\partial(\frac{\Psi_{Mjt}}{\Psi_{Wjt}})}{\partial HSR}$.

Table 14 summarizes three parameters used in the quantitative estimation. The choice of elasticity of substitution in the production function between male- and female-intensive inputs (σ) and the labor share in Cobb-Douglas Function in each male- and female-intensive sectors (β)

are borrowed from the Bank of Korea estimates (Bae (2014)). CDF parameter of labor force participation cost (ι) is calibrated. Details of the calibration are provided in Appendix A7.

Table 14 about here

Table 15 presents the results of the quantitative estimation of the impact of HSR, solving Equations 8 and 9 given the set of parameters in Table 14. HSR reduces labor force participation costs for women in core areas by 4.74% and in non-core areas by 16.71%. In contrast, the relative productivity of male-intensive sectors decreased by 30.59% and 10.63%, respectively, in core and non-core areas. The estimation results show that HSR mainly benefits women in core areas through increased relative productivity in female-intensive sectors. Women in non-core areas benefit more from the decreases in labor force participation costs, with the expansion of HSR.

Table 15 about here

10 Robustness Check

10.1 Subset Analysis

In this section, robustness checks are performed to validate the results found in the previous sections. One of the main concerns is if outliers may drive the empirical results. To address this, the same sets of regression in Section 7.1 with subsets of the sample are used: (1) without districts designated as a special district by the government (i.e., Sejong City); (2) without districts located in big cities; (3) without districts located close to the North Korean border. Neither of the subsample analyses change the main findings of the results in Section 7.1.

10.1.1 Without Sejong cities

First, the subsample without Sejong City, an autonomous city in South Korea, is considered. Since 2012, the South Korean government has relocated numerous ministries and agencies to Sejong. If this new city's influence drove the redistribution of local economic outcomes, this could compromise our primary regression's causal relationship. As we can observe from Table 16, the estimates barely changed in the estimation without Sejong city, which confirms the validity of the main findings.

Table 16 about here

10.1.2 Without Districts in Big Cities

Next, Table 17 contains a set of regressions without the following "big" cities: Seoul, Busan, Daejeon, Daegu, and Gwangju. All the results are quantitatively similar to our main regressions, except for the gender gap in employment (the third column in the upper panel). The impact of HSR on the reduction in the gender employment gap in non-core areas dissipate if we remove the big cities in the non-core areas from the variables. This means that the decline in the gender gap in non-core areas mainly occurs in the highly industrialized areas, whereas reduction was not significant in rural areas.

Table 17 about here

10.1.3 Without Districts near North Korean Border

Finally, Table 18 presents the subsample analysis without districts closer to the North Korean border, which could follow a different trajectory in growth due to the political uncertainty. The regression results are qualitatively and quantitatively close to the main results.

Table 18 about here

10.2 Nevo and Rosen (2012)'s Identification with Imperfect Instruments

Using old railroad stations as an instrument strategy is widely accepted in the urban economics literature, but this approach has limitations. For instance, if the path dependency in local development exists, even if Japanese construction of the railroad did not target the local economy growth after 100 years, the existence of old railroad stations could contribute to today's growth of the districts. To address this issue, in this section, Nevo and Rosen (2012)'s identification with imperfect instrument idea was used as a robustness check of the empirical estimates using instrumental variables.

Nevo and Rosen (2012) allow the instrumental variable to be correlated with the error term; however, they assume the correlation between the instrumental variable and the error term has the same sign as the correlation between the endogenous regressor and the error term. Moreover, the instrumental variable is assumed to be less correlated with the error term than is the endogenous regressor. These assumptions provide the bounds for the parameters using the instrumental variable instead of the exact estimates. Qualitatively, the bounds estimated using Nevo and Rosen (2012) are similar to our main regression results in Section 7.1.

11 Conclusion

This study investigates the effect of HSR on the redistribution of economic activities and the gender differences in labor market outcomes. Considering construction of the old railroad in the Japanese invasion era as an instrumental variable, this study demonstrated that population and employment move from core areas to non-core areas due to the construction of the HSR in South Korea. Women benefit more from HSR expansion than men because the local service sector wherein women mainly work benefits more from reduction in travel time with HSR as it reduces the cost of moving people rather than that of goods. Moreover, with changes in endogenous local amenities with HSR, especially changes in education amenities, women benefit more from HSR as it potentially reduces the burden of childcare and other domestic burdens, which is not a significant concern for men.

With the construction of the old railroad in the Japanese invasion era as an instrumental variable, this paper empirically shows that HSR narrows the gender employment gap in both core areas (i.e., districts in the Seoul metropolitan areas) non-core areas. The wage disparities between men and women in core areas decrease with HSR.

The mechanisms are structurally decomposed using a spatial general equilibrium model. The HSR's impact on the gender employment and wage gaps can be decomposed into the labor demand and the labor supply channels in the model. The quantitative decomposition shows that overall, HSR increases the labor demand of female-intensive sectors and decreases women's labor participation costs. Specifically, in core areas, the labor demand impacts are more significant than the labor supply impacts, whereas, in non-core areas, the labor supply impacts are more distinctive.

Finally, the model-predicted mechanisms were empirically explored. Women benefit more from HSR expansion than men, as the local service sector in which women mainly work benefits more from reducing travel time with HSR than the sectors where men mostly work. Additionally, the improvement in local amenities, particularly in education and childcare facilities, reduces women's childcare burden and encourages women to join the labor force.

This paper contributes to the literature by focusing on the impact of transportation infrastructure on the gender gap in labor market outcomes, which there exists limited prior evidence. Furthermore, this study extends the literature by examining how improved transportation technology could impact both the demand and supply side of the gender-segmented labor market. This paper's findings shed light on the importance of understanding the heterogeneous impacts of infrastructure investments across different demographic groups, which is an essential consideration, given the considerable spending on the transportation infrastructure.

This study is not without limitations. First, the study does not cover much of the overall welfare for men and women. Moreover, it tests how the HSR affects the economic opportunity for women somewhat indirectly; however, did not provide definite evidence of why women's labor force participation costs would be decreased. Future studies related to this topic could address these questions.

References

Ahlfeldt, G. M. and Feddersen, A. (2018). From periphery to core: measuring agglomeration effects using high-speed rail. *Journal of Economic Geography*, 18(2):355–390.

Aksoy, C. G., Berkay, O., and Philip, J. (2019). Robots and the gender pay gap: Evidence from europe. In *2nd IZA/CREA Workshop: Exploring the Future of Work*, volume 5.

Bae, B. (2014). Report on construction of new bok-dsge model for economic forecasting and policy analysis.

Baek, J. and Park, W. (2015). The impact of improved passenger transport system on manufacturing plant productivity. *KDI School of Pub Policy & Management Paper*, (15-02).

Bartik, T. J. (1991). Who benefits from state and local economic development policies?

Baum-Snow, N. (2007). Did highways cause suburbanization? *The Quarterly Journal of Economics*.

Baum-Snow, N., Henderson, J. V., Turner, M. A., Zhang, Q., and Brandt, L. (2020). Does investment in national highways help or hurt hinterland city growth? *Journal of Urban Economics*, 115:103124.

Black, D. A., Kolesnikova, N., and Taylor, L. J. (2014). Why do so few women work in new york (and so many in minneapolis)? labor supply of married women across us cities. *Journal of Urban Economics*, 79:59–71.

Blau, F. D. and Kahn, L. M. (2017). The gender wage gap: Extent, trends, and explanations. *Journal of Economic Literature*, 55(3):789–865.

Charnoz, P., Lelarge, C., and Trevien, C. (2018). Communication costs and the internal organisation of multi-plant businesses: evidence from the impact of the french high-speed rail. *The Economic Journal*, 128(610):949–994.

Chauvin, J. P. (2017). Gender-segmented labor markets and the effects of local demand shocks. *Job Market Paper, Harvard University.*

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

Donaldson, D. (2018). Railroads of the raj: Estimating the impact of transportation infrastructure. *American Economic Review*, 108(4-5):899–934.

Donaldson, D. and Hornbeck, R. (2016). Railroads and american economic growth: A "market access" approach. *The Quarterly Journal of Economics*, 131(2):799–858.

Dong, X., Zheng, S., and Kahn, M. E. (2020). The role of transportation speed in facilitating high skilled teamwork across cities. *Journal of Urban Economics*, 115:103212.

Duranton, G., Morrow, P. M., and Turner, M. A. (2014). Roads and trade: Evidence from the us. *Review of Economic Studies*, 81(2):681–724.

Faber, B. (2014). Trade integration, market size, and industrialization: evidence from china's national trunk highway system. *Review of Economic Studies*, 81(3):1046–1070.

Farré, L., Jofre-Monseny, J., and Torrecillas, J. (2020). Commuting time and the gender gap in labor market participation.

Garcia-López, M.-À., Holl, A., and Viladecans-Marsal, E. (2015). Suburbanization and highways in spain when the romans and the bourbons still shape its cities. *Journal of Urban Economics*, 85:52–67.

Glaeser, E. L. et al. (2008). Cities, agglomeration, and spatial equilibrium.

Goldin, C. and Katz, L. F. (2016). A most egalitarian profession: pharmacy and the evolution of a family-friendly occupation. *Journal of Labor Economics*, 34(3):705–746.

Gorback, C. (2020). *Essays in Urban and Real Estate Economics*. PhD thesis, University of Pennsylvania.

Greenwood, J., Guner, N., Kocharkov, G., and Santos, C. (2016). Technology and the changing family: A unified model of marriage, divorce, educational attainment, and married female labor-force participation. *American Economic Journal: Macroeconomics*, 8(1):1–41.

Greenwood, J., Seshadri, A., and Yorukoglu, M. (2005). Engines of liberation. *The Review of Economic Studies*, 72(1):109–133.

Heuermann, D. F. and Schmieder, J. F. (2019). The effect of infrastructure on worker mobility: evidence from high-speed rail expansion in germany. *Journal of economic geography*, 19(2):335–372.

Kawabata, M. and Abe, Y. (2018). Intra-metropolitan spatial patterns of female labor force participation and commute times in tokyo. *Regional Science and Urban Economics*, 68:291–303.

Le Barbanchon, T., Rathelot, R., and Roulet, A. (2020). Gender differences in job search: Trading off commute against wage.

Lin, Y. (2017). Travel costs and urban specialization patterns: Evidence from china's high speed railway system. *Journal of Urban Economics*, 98:98–123.

Liu, S. and Su, Y. (2020). The geography of jobs and the gender wage gap. *Available at SSRN.*

Maurer-Fazio, M., Connelly, R., Chen, L., and Tang, L. (2011). Childcare, eldercare, and labor force participation of married women in urban china, 1982–2000. *Journal of Human Resources*, 46(2):261–294.

Moretti, E. (2010). Local labor markets. Technical report, National Bureau of Economic Research.

Morten, M. and Oliveira, J. (2016). Paving the way to development: Costly migration and labor market integration. Technical report, National Bureau of Economic Research.

Nevo, A. and Rosen, A. M. (2012). Identification with imperfect instruments. *Review of Economics and Statistics*, 94(3):659–671.

Qin, Y. (2017). 'no county left behind?'the distributional impact of high-speed rail upgrades in china. *Journal of Economic Geography*, 17(3):489–520.

Redding, S. J. and Turner, M. A. (2014). Transportation Costs and the Spatial Organization of Economic Activity. *CEP Discussion Paper*, (No. 1277).

Roback, J. (1982). Wages, rents, and the quality of life. *Journal of political Economy*, 90(6):1257–1278.

Rosenthal, S. S. and Strange, W. C. (2012). Female entrepreneurship, agglomeration, and a new spatial mismatch. *Review of Economics and Statistics*, 94(3):764–788.

Severen, C. (2018). Commuting, labor, and housing market effects of mass transportation: Welfare and identification.

Stock, J. H. and Yogo, M. (2002). Testing for weak instruments in linear iv regression. Technical report, National Bureau of Economic Research.

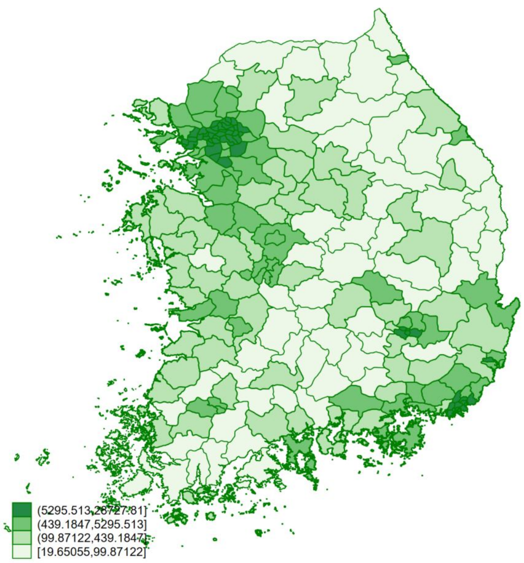
Tsivanidis, N. (2018). The aggregate and distributional effects of urban transit infrastructure: Evidence from bogotá's transmilenio. *Job Market Paper*.

Vidart, D. (2020). Human capital, female employment, and electricity: Evidence from the early 20th century united states. *Unpublished Working Paper, UCSD*.

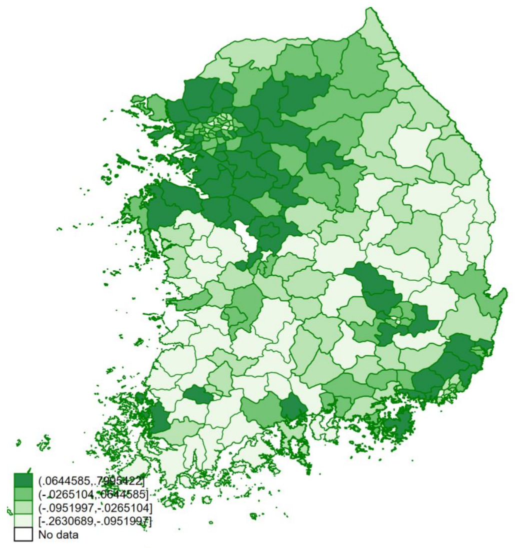
Zheng, S. and Kahn, M. E. (2013). China's bullet trains facilitate market integration and mitigate the cost of megacity growth. *Proceedings of the National Academy of Sciences*, 110(14):E1248–E1253.

Figure 1: Spatial Distribution of Population and Employment in South Korea

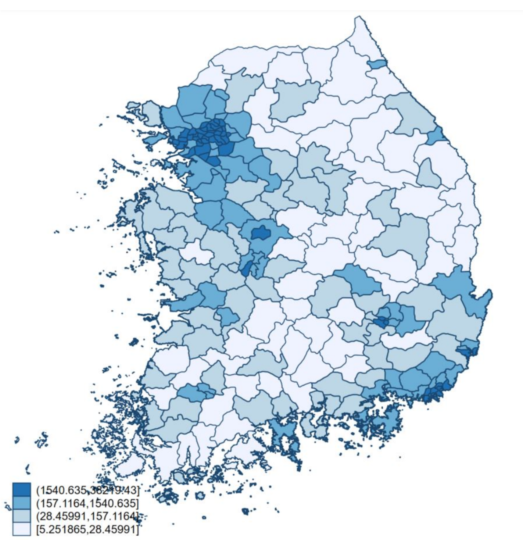
(a) *Population* (2010)



(b) *Population Growth* (2003-2012)



(c) *Employment* (2010)



(d) *Employment Growth* (2003-2012)

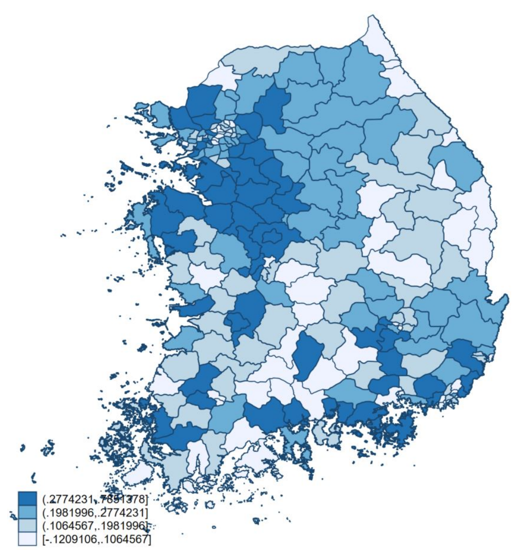
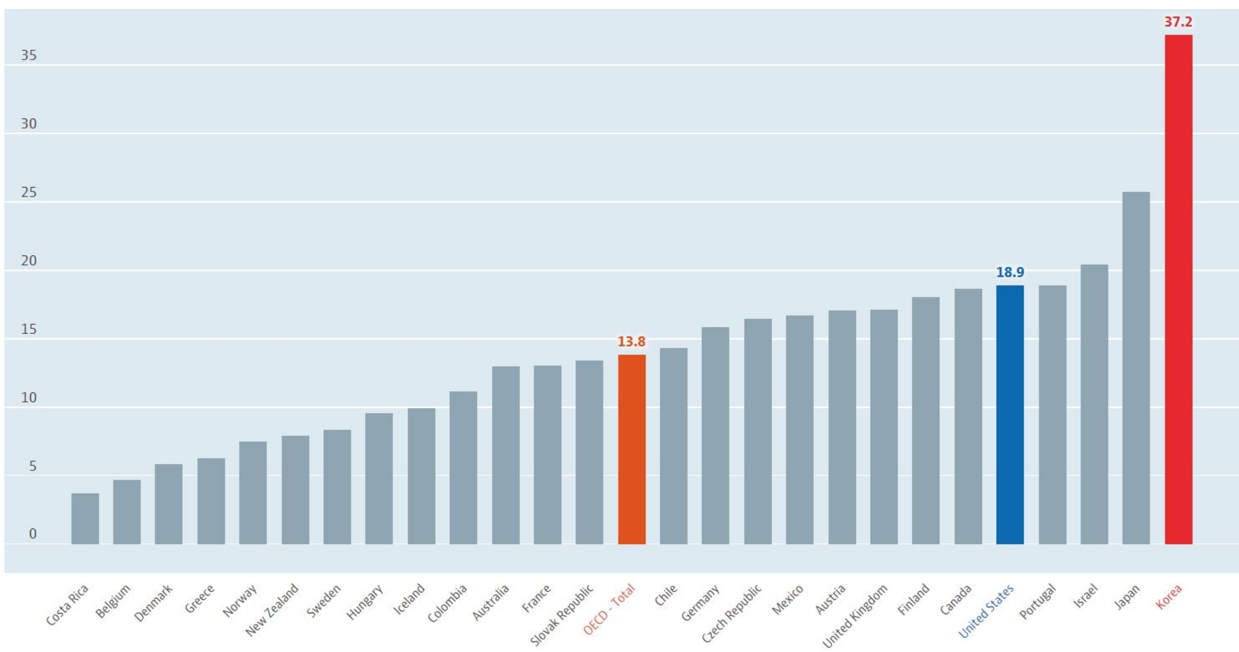
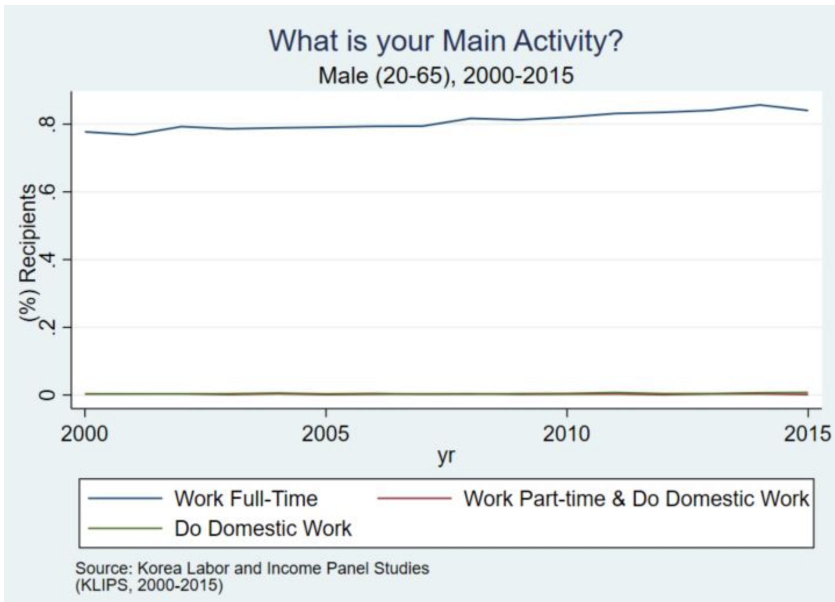


Figure 2: Cross-Country Gender Wage Gap (year=2015)



Note: The gender wage gap is defined as the difference between median earnings of men and women relative to median earnings of men. Data refer to full-time employees on the one hand and to self-employed on the other.
Source: OECD (2020), Gender wage gap (indicator)

Figure 3: Gender Differences in Time Use across Years



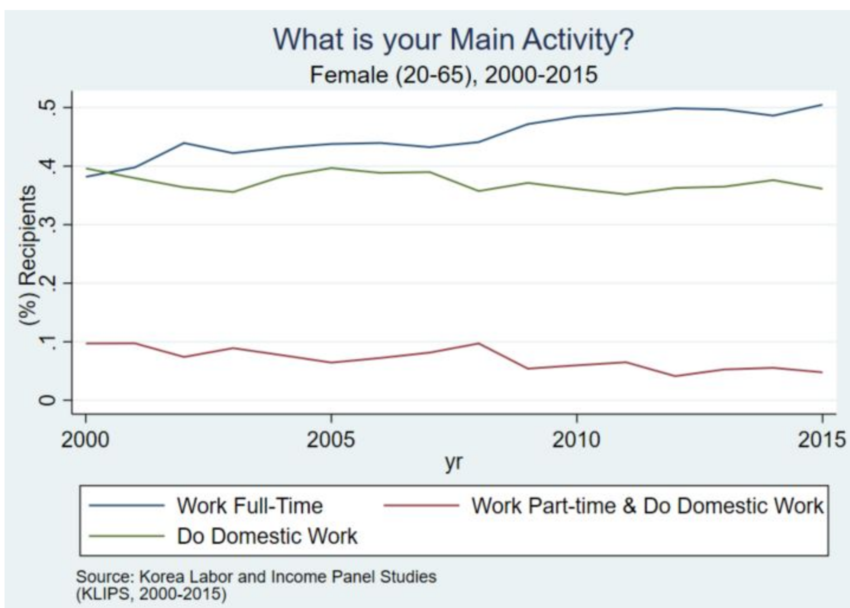


Figure 4: Wage Differences between Men and Women across Years

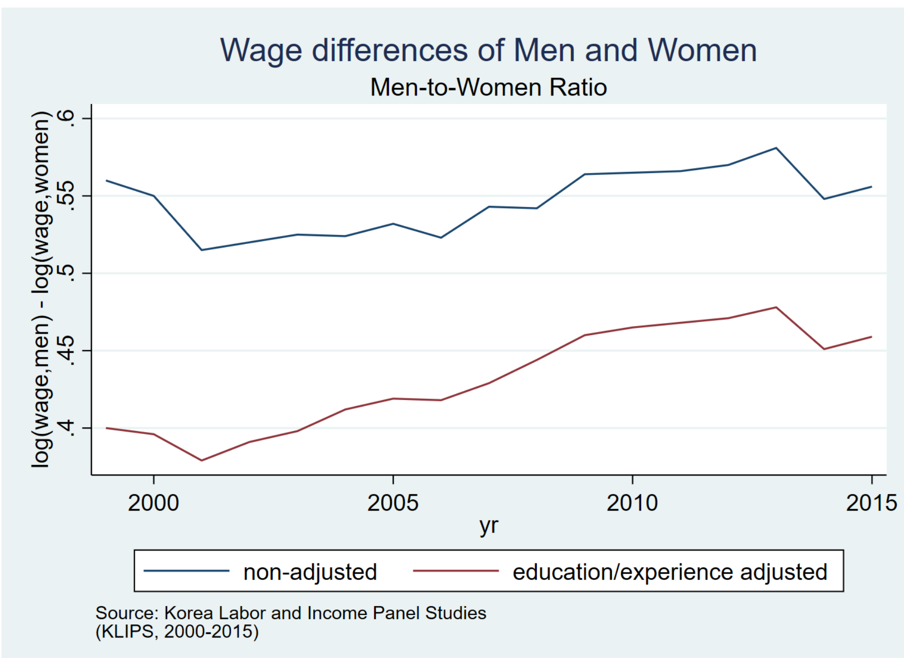
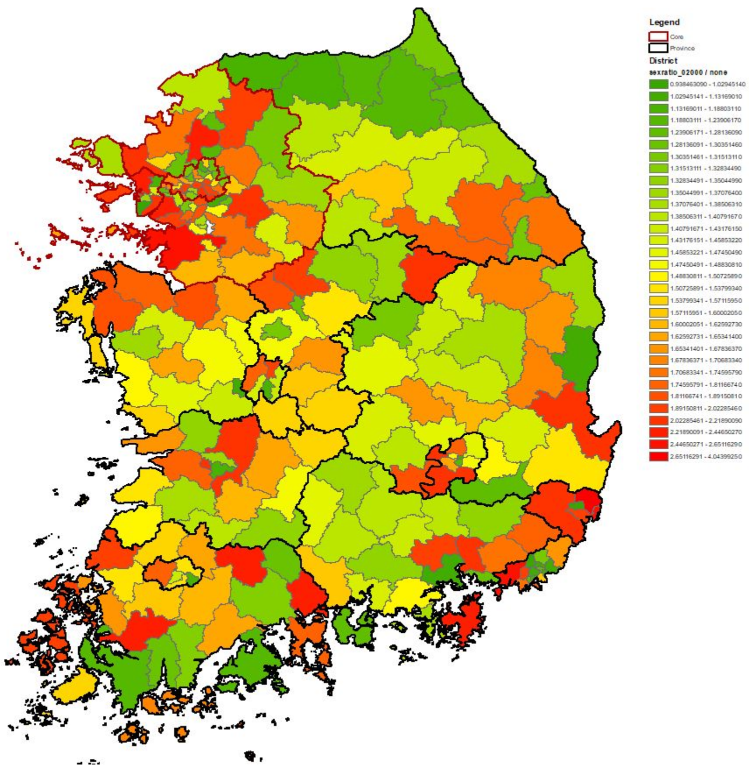
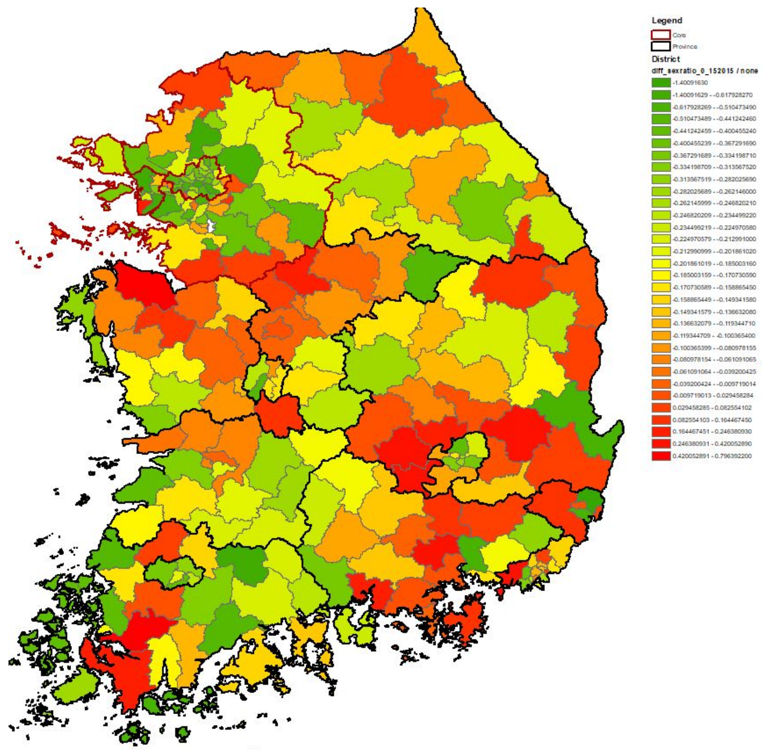


Figure 5: Spatial Distribution of Employment Sex Ratio

(a) *Sex Ratio* (2000)



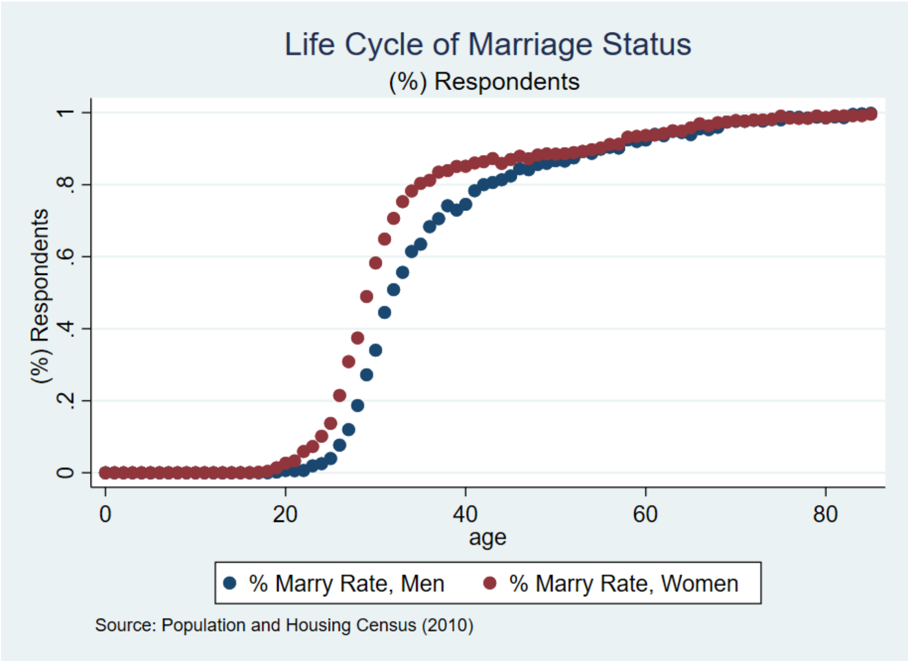
(b) *Changes in Sex Ratio* (2000-2015)



Note: Sex ratio of employment is defined as local male employment over female employment. Source: Census on Establishment (Statistics Korea)

Figure 6: Life Cycle of Work and Marriage Status

(a) (%) Married



(b) (%) Working

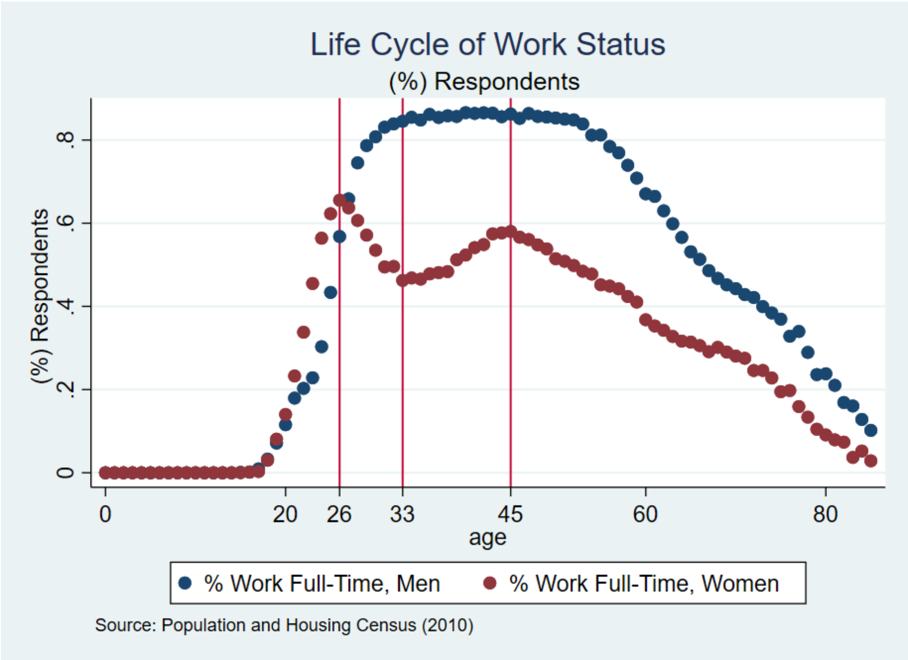
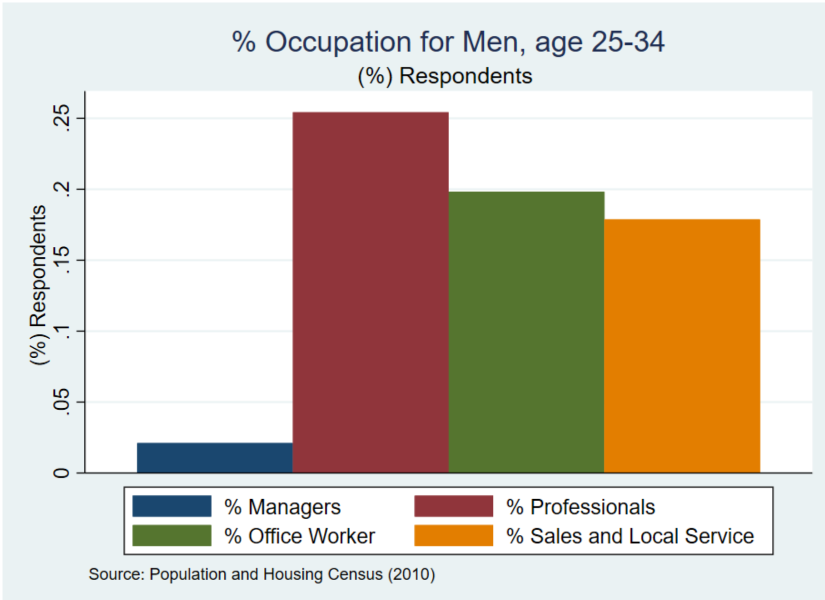
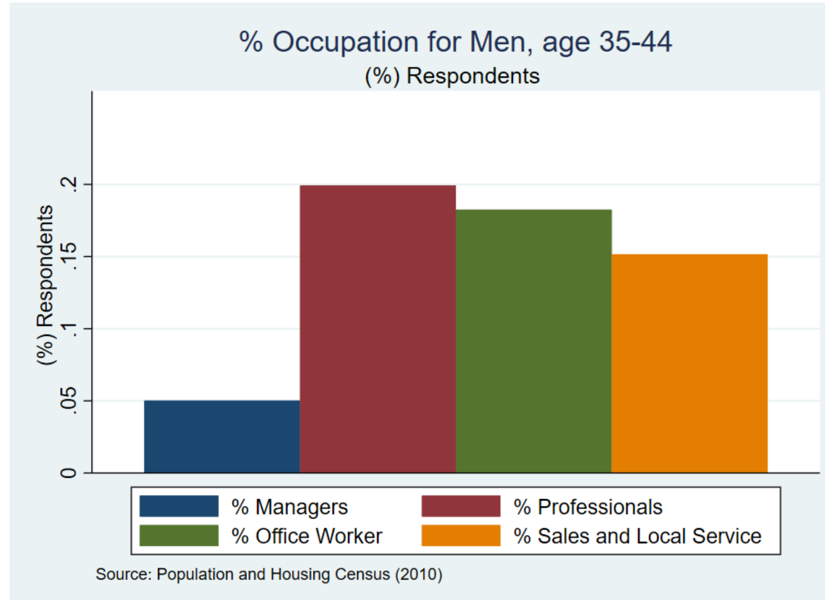


Figure 7: Occupation Composition for Gender and Age Group

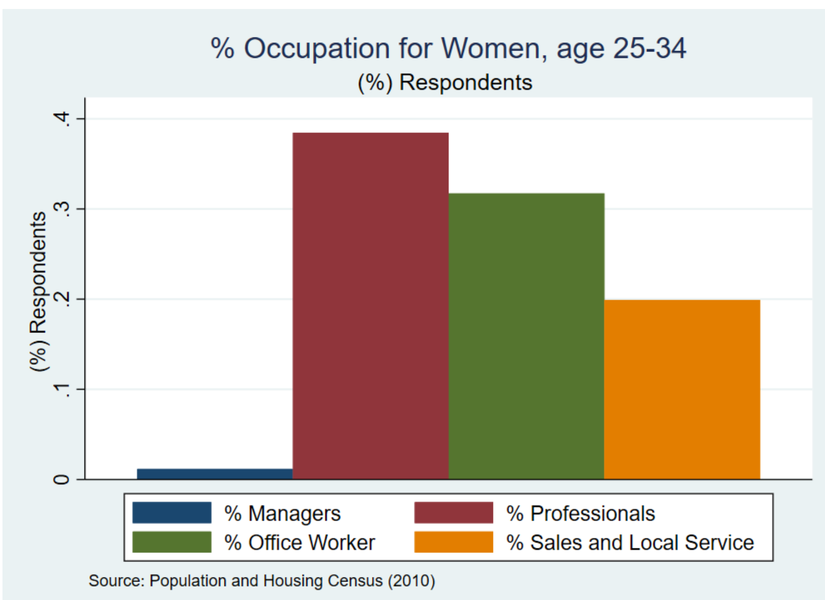
(a) *Men, age 25-34*



(b) *Men, age 35-44*



(c) *Women, age 25-34*



(d) *Women, age 35-44*

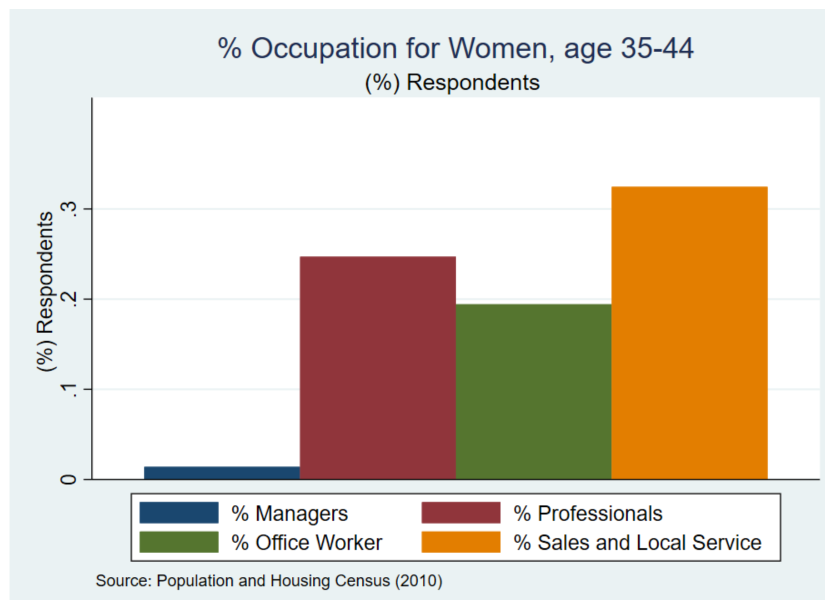
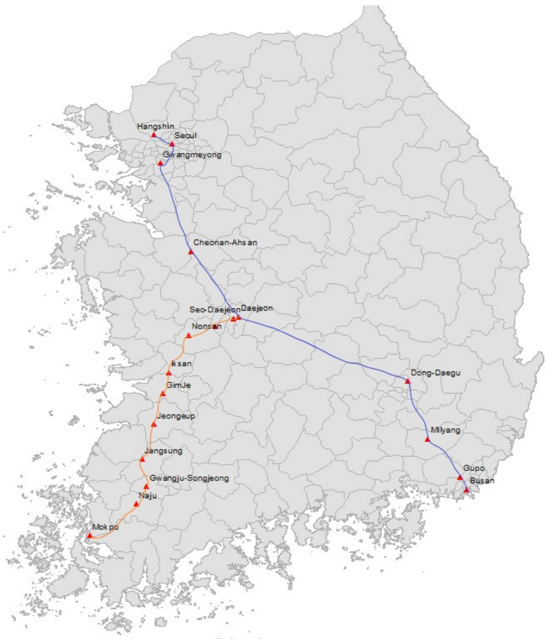
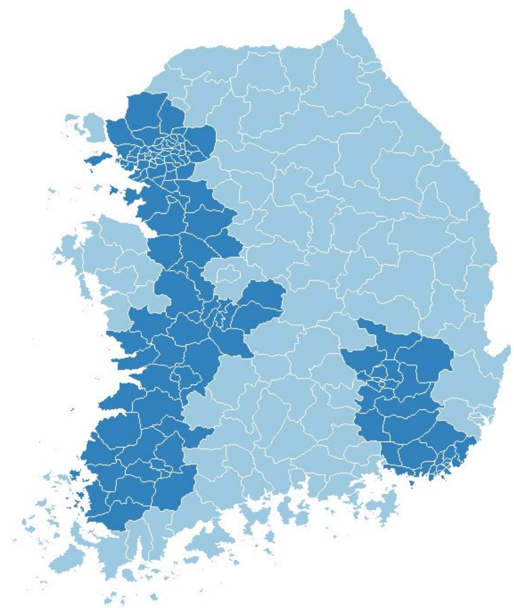


Figure 8: KTX Network in South Korea

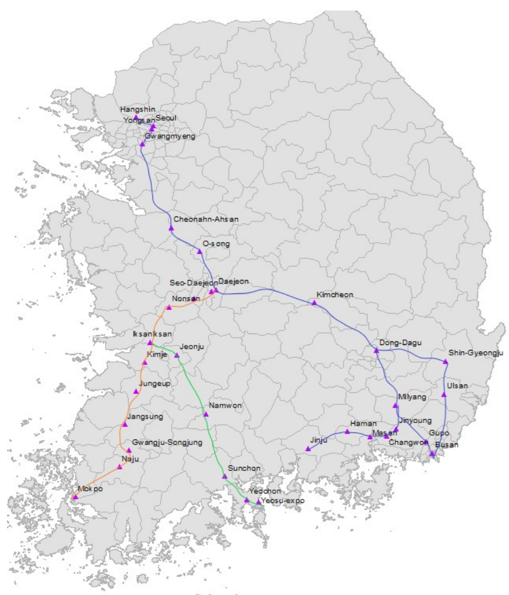
(a) *Phase 1* KTX Station (2004)



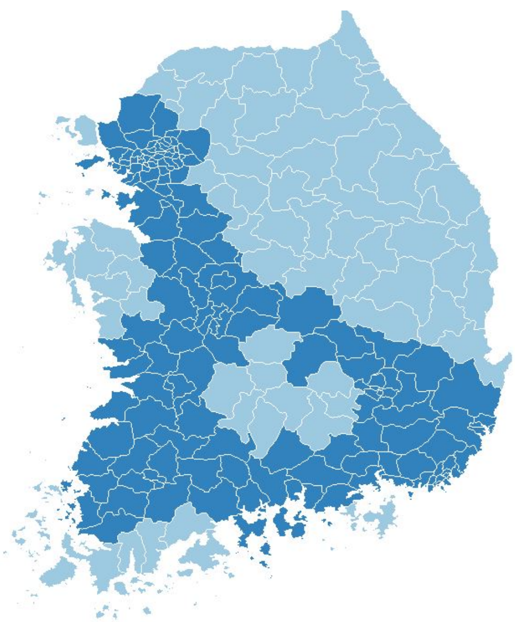
(b) *Phase 1* Area Covered by KTX (2004)



(c) *Phase 2* KTX Station (2012)



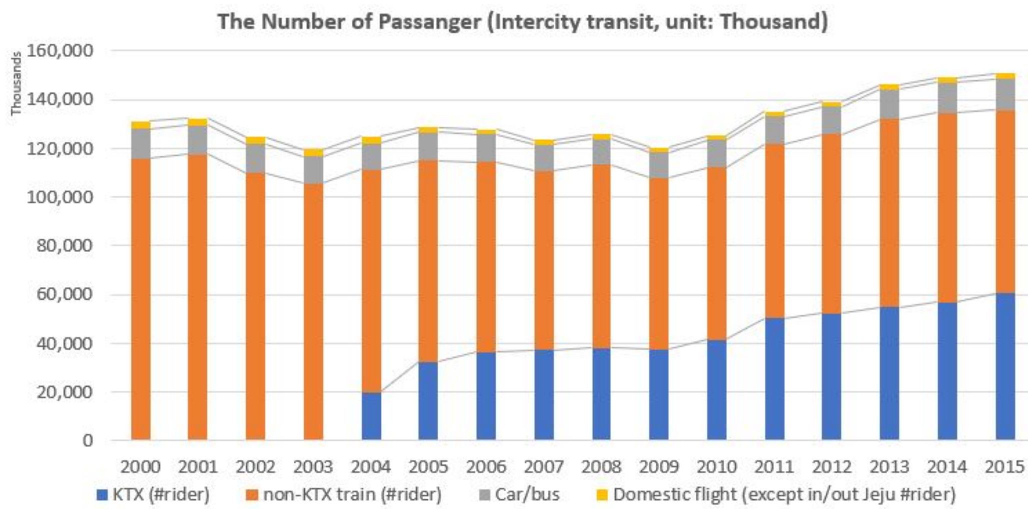
(d) *Phase 2* Area Covered by KTX (2012)



Note: Area covered by KTX are districts with a centroid distanced less than km from any KTX stations

Figure 9: Annual Ridership across Different Transportation Modes

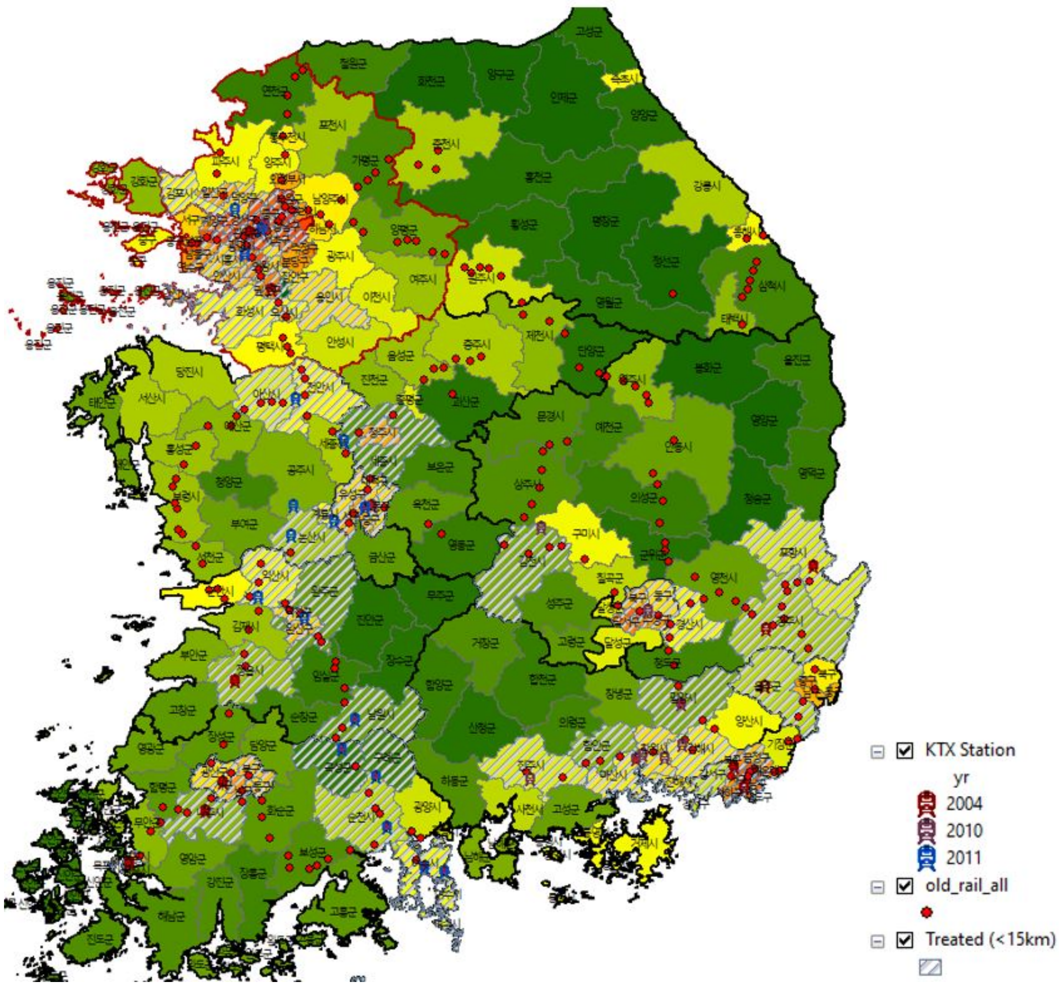
(a)



(b)

Source: Korea Transportation Database (KTDB)

Figure 10: KTX Network and Old Railroad Stations Constructed during Japanese Colonial Era



Note: Color represents the population density of districts in 2010. Districts within the thick red lines are defined as core areas, districts in Seoul metropolitan areas, which are districts of Incheon, Seoul, or Gyeonggi provinces. Shaded districts are 'treated' districts, where the population-weighted centroid of districts is within 15km of any high-speed rail stations; and zero, otherwise. Red dots are the location of old railroad stations constructed during the Japanese colonial era.

Table 1: Descriptive Statistics of Demographics

Panel A. Share of Working Population and Single

	Working (%)		Single (%)	
	Female	Male	Female	Male
Core	39.72%	75.68%	12.40%	12.96%
Non-core	46.96%	75.53%	6.92%	7.50%

'Working' includes any types of employment(e.g. formal, informal). Statistics are share of (fe)male, who work (Column (1), Column (2)); and who are not married, i.e. single (Column (3), Column (4)). Statistics are mean in 2005, over age 25. Census on Population and Household, Statistics Korea.

Panel B. Gender Gap in Labor Market Outcomes

	Employment Gap (male-to-female,%)	Wage Gap (male-to-female,%)
Core	40.91%	38.8%
Non-core	38.63%	40.3%

Employment gap is defined as local male employment over female employment. Wage gap is local male-to-female wage ratio. The statistics are 3 year averages (2000-2003). Census on Establishment, Korea Labor Income Panel Survey (KLIPS).

Table 2: Model Prediction: Impact of HSR

Panel A. Impact on the Gender Employment Gap (Male/Female)

	Female Labor Participation Cost $(1 + T_{jt})$		
Relative Productivity $(\frac{\Psi_{Mjt}}{\Psi_{Wjt}})$	**Decrease**	**No Changes**	**Increase**
Decrease	(-)	(-)	(?)
No Changes	(-)	(.)	(+)
Increase	(?)	(+)	(+)

Panel B. Impact on the Gender Wage Gap (Male/Female)

	Female Labor Participation Cost $(1 + T_{jt})$		
Relative Productivity $(\frac{\Psi_{Mjt}}{\Psi_{Wjt}})$	**Decrease**	**No Changes**	**Increase**
Decrease	(?)	(-)	(-)
No Changes	(+)	(.)	(-)
Increase	(+)	(+)	(?)

Table 3: Summary Statistics

Variable	(1) Mean	(2) Std Dev	(3) Min	(4) Median	(5) Max
log(Employment)	10.639	1.065	7.954	10.789	13.475
log(Establishment)	9.154	0.939	6.811	9.253	11.206
Sex ratio (Employment)	1.460	0.364	0.764	1.383	4.044
Sex ratio (Population)	1.006	0.041	0.898	1.005	1.308
log(Landprice)	87.267	9.316	46.527	90.003	105.563
log(Wage, female)	-0.399	0.254	-2.377	-0.382	0.538
log(Wage, male)	0.038	0.240	-1.664	0.038	1.184
log(Wage, male)- log(Wage, female)	0.437	0.240	-0.596	0.434	2.201
log(Employment): Transportation	7.613	1.286	4.477	7.767	10.815
log(Employment): Construction	7.720	1.095	5.182	7.637	11.136
log(Employment): Public Admin	7.578	0.719	5.011	7.540	10.325
log(Employment): Manufacturing	8.817	1.403	4.382	8.802	12.440
log(Employment): Retail Store	7.024	0.932	4.407	7.064	9.616
log(Employment): Education	8.067	1.078	5.357	8.137	11.140
log(Employment): Medical Service	7.217	1.240	4.060	7.286	10.386
log(Employment): Restaurant	8.545	0.967	6.174	8.684	11.107
Establishment per population: Retail	0.003	0.002	0.001	0.003	0.011
Establishment per population: Education	0.003	0.001	0.001	0.002	0.012
Establishment per population: Medical Service	0.001	0.001	0.000	0.001	0.005
Establishment per population: Restaurant	0.016	0.008	0.005	0.014	0.059
Education Workers Per Kids ($age < 14$)	9.417	2.482	14.295	87.968	16.707

Notes: All the statistics are the district-level, from 2000 to 2015.

Table 4: First Stage of Instrumental Variable Regression

	(1)	(2)	(3)
		First Stage	
VARIABLES	D=1: Treat	D=1: Treat*Core	D=1: Treat*Non-core
Old Railroad Station	0.564***		
	(0.63)		
**Non-core areas*		0.000	0.542***
		(0.042)	(0.054)
**Core areas*		0.687***	-0.000
		(0.100)	(0.128)
State Fixed Effect	✓	✓	✓
Observations	238	238	238
R-squared	0.354	0.612	0.485
F-Statistics	15.69	39.87	23.93

Notes: *** p<0.01, ** p<0.05, * p<0.1. Column (1)-(3) reports the first stage of the instrumental variable regression. Column (1) is the first stage for the regression without heterogeneous treatment effects and column (2), (3) are with heterogeneous treatment effects. Old Railraod Station variable is equal to one, if the population-weighted centroid of districts is within 15km of any Japanese constructed train stations; or zero, otherwise. Core areas are defined as districts in Seoul metropolitan areas, which are districts of Incheon, Seoul, or Gyeonggi provinces. Non-core areas are all other districts.

Table 5: Impact of HSR on Local Population, Employment and the number of Establishments (Two-way Fixed Effects Model)

log of	(1) Population	(2) Employment	(3) Establishment	(4) Population	(5) Employment	(6) Establishment
Treat	0.033*	-0.002	0.022			
	(0.017)	(0.020)	(0.020)			
**Non-core areas*				0.073***	0.039*	0.066***
				(0.018)	(0.021)	(0.020)
**Core areas*				-0.056	-0.097**	-0.078*
				(0.036)	(0.042)	(0.042)
Year FE	✓	✓	✓	✓	✓	✓
District FE	✓	✓	✓	✓	✓	✓
Province-Year FE	✓	✓	✓	✓	✓	✓
Observations	3,806	3,806	3,806	3,806	3,806	3,806
R-squared	0.994	0.993	0.992	0.994	0.993	0.992
N. Districts	238	238	238	238	238	238

Notes: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Standard errors are in parentheses. The coefficients are the regression results of instrumental variable estimation in Section 6.1. Dependent variable of column (1), (4) is log of population, column (2), (5), is log of employment, and column (3), (6) is log of the number of establishments respectively. Treat variable is equal to one, if the population-weighted centroid of districts is within 15km of any high-speed rail stations; or zero, otherwise. Core areas are defined as districts in Seoul metropolitan areas, which are districts of Incheon, Seoul, or Gyeonggi provinces. Non-core areas are all other districts. Standard errors are clustered at district-level.

Table 6: Impact of HSR on Local Population, Employment and the number of Establishments (IV)

	(1)	(2)	(3)	(4)	(5)	(6)
log of	Population	Employment	Establishment	Population	Employment	Establishment
Treat	0.114	0.032	0.055			
	(0.087)	(0.120)	(0.125)			
**Non-core areas*				0.188***	0.132**	0.161***
				(0.045)	(0.052)	(0.047)
**Core areas*				-0.207**	-0.407***	-0.410***
				(0.085)	(0.086)	(0.090)
Province FE	✓	✓	✓	✓	✓	✓
1st stage F-stat	15.69	15.69	15.69			
(Non-core)				39.87	39.87	39.87
(Core)				23.93	23.93	23.93
Observations	238	238	238	238	238	238
R-squared	0.161	0.092	0.094	0.191	0.043	0.018

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors are in parentheses. The coefficients are the regression results of instrumental variable estimation in Section 6.2. Dependent variable of column (1), (4) is log of population, column (2), (5), is log of employment, and column (3), (6) is log of the number of establishments respectively. Treat variable is equal to one, if the population-weighted centroid of districts is within 15km of any high-speed rail stations; or zero, otherwise. Core areas are defined as districts in Seoul metropolitan areas, which are districts of Incheon, Seoul, or Gyeonggi provinces. Non-core areas are all other districts. Standard errors are clustered at district-level.

Table 7: Impact of HSR on Population, Employment and Employment Rate across Gender (*IV*)

	(1)	(2)	(3)	(4)	(5)	(6)
	Population			*Employment*		
	Male	Female	Sex Ratio	Male	Female	Sex Ratio
Treat						
**Non-core*	0.177***	0.200***	-0.025*	0.120*	0.209***	-0.162**
	(0.046)	(0.045)	(0.013)	(0.060)	(0.059)	(0.068)
**Core*	-0.243**	-0.168*	-0.083	-0.438***	-0.295***	-0.199***
	(0.095)	(0.084)	(0.067)	(0.070)	(0.090)	(0.055)
Province FE	✓	✓	✓	✓	✓	✓
1st stage F-stat						
(Non-core)	39.87	39.87	39.87	39.87	39.87	39.87
(Core)	23.93	23.93	23.93	23.93	23.93	23.93
Observations	238	238	238	237	237	237
R-squared	0.178	0.202	0.014	0.034	0.094	0.095

Notes: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Standard errors are in parentheses. The coefficients are the regression results of instrumental variable estimation in Section 6.2. Dependent variable of column (1), (2) is log of population of male and female, and column (3) is sex ratio of population (male population/ female population). Column (4), (5), is log of employment of male and female, and column (6) is sex ratio of employment (male employment/ female employment). Treat variable is equal to one, if the population-weighted centroid of districts is within 15km of any high-speed rail stations; and zero, otherwise. Core areas are defined as districts in Seoul metropolitan areas, which are districts of Incheon, Seoul, or Gyeonggi provinces. Non-core areas are all other districts. Standard errors are clustered at province-level.

Table 8: Impact of HSR on Wages, and Gender Wage Gaps (*IV*)

	(1)	(2)	(3)
		IV estimates	
	logwage (male)	logwage (female)	wage gap (male/female)
Treat			
**Non-core areas*	0.072	0.128	0.003
	(0.257)	(0.175)	(0.096)
**Core areas*	0.046	0.203**	-0.158***
	(0.100)	(0.085)	(0.017)
Province FE	✓	✓	✓
1st stage F-stat			
(Non-core)	39.87	39.87	39.87
(Core)	23.93	23.93	23.93
Observations	179	166	162
R-squared	0.046	0.011	0.057

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors are in parentheses. The coefficients are the regression results of instrumental variable estimation in Section 6.2. Dependent variable of column (1), (2), is log of wage of men and women, respectively, and column (3) is gender wage gap, defined as local wage of men relative to local wage of women. Treat variable is equal to one, if the population-weighted centroid of districts is within 15km of any high-speed rail stations; and zero, otherwise. Core areas are defined as districts in Seoul metropolitan areas, which are districts of Incheon, Seoul, or Gyeonggi provinces. Non-core areas are all other districts. Standard errors are clustered at province-level.

Table 9: Summaries of Reduced Form Findings

	Population (Employment)	Gender Gap (male/female)	
		Employment	Wage
Core	(-)	(-)	(-)
Non-core	(+)	(-)	(.)

Table 10: Mechanism 1.(1). Effect of HSR on Sectoral Employment

Panel A. Male Intensive Sector

	(1)	(2)	(3)	(4)
	Log of Employment			
	(Transportation)	(Construction)	(Public Admin)	(Manufacturing)
Treat				
**Non-Seoul*	0.241	0.251*	-0.016	-0.316**
	(0.137)	(0.134)	(0.114)	(0.136)
**Seoul*	-0.735**	-0.490**	-0.189***	-0.925***
	(0.310)	(0.205)	(0.062)	(0.269)
Sex Ratio(2000)	10.09	7.56	4.40	2.36
Observations	237	237	237	237

Panel B. Female Intensive Sector

	(5)	(6)	(7)	(8)
	Log of Employment			
	(Retail)	(Education)	(Medical Service)	(Restaurant)
Treat				
**Non-Seoul*	0.251***	0.333***	0.474***	0.205***
	(0.060)	(0.071)	(0.085)	(0.061)
**Seoul*	-0.328*	-0.048	-0.090	-0.327***
	(0.172)	(0.087)	(0.182)	(0.025)
Sex Ratio(2000)	0.88	0.76	0.54	0.50
Observations	237	237	236	237

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors are in parentheses. The coefficients are the second stage regression results of instrumental variable estimation in Section 6.2. Dependent variable of column (1) - (8) is log of employment in each sector. Core areas are defined as districts in Seoul metropolitan areas, which are districts of Incheon, Seoul, or Gyeonggi provinces. Non-core areas are all other districts. Standard errors are clustered at province-level.

Table 11: Mechanism 2. HSR's Impact on Endogeneous Amenities in Non-core

	(1)	(2)	(3)	(4)	(5)
SECTOR	Retail Stores	Medical Service	Restaurant	Education	Education
VARIABLE	Establishment per Residents				Workers per Child ($age < 13$)
Treat					
**Non-Seoul*	0.219***	0.161***	0.052	0.249***	0.016*
	(0.059)	(0.050)	(0.049)	(0.055)	(0.009)
**Seoul*	0.065	0.118	-0.054	0.131	0.036
	(0.129)	(0.151)	(0.181)	(0.179)	(0.037)
Province FE	✓	✓	✓	✓	✓
Province FE	✓	✓	✓	✓	✓
1st stage F-stat					
(Non-core)	39.87	39.87	39.87	39.87	39.87
(Core)	23.93	23.93	23.93	23.93	23.93
Observations	237	236	236	237	237

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors are in parentheses. The coefficients are the second stage regression results of instrumental variable estimation in Section 6.2. Dependent variable of column (1), (2), (3), (4) is log of the number of retail (KSIC G471: department stores, supermarkets, grocery stores, etc); education (KSIC P: Private and Public education institutes); medical service(KSIC Q86: hospitals and all types of medical services); and restaurant (KSIC I56) establishments per resident respectively, following Diamond (2016). Dependent variable of column (5) is number of workers in education sector per child below age of 13. Core areas are defined as districts in Seoul metropolitan areas, which are districts of Incheon, Seoul, or Gyeonggi provinces. Non-core areas are all other districts. Standard errors are clustered at province-level.

Table 12: Effect of Travel Time between Origin and Destination Districts on Migration: Single

VARIABLES	(1)	(2)	(3)
	log of Number of Migrants		
	(Single, all)	(Single, Male)	(Single, Female)
Panel A.			
log(Travel Time)	-0.050***	-0.050***	-0.045***
	(0.013)	(0.013)	(0.013)
Observations	849660	849660	849660
R2	0.930	0.912	0.898
Panel B.			
log (Travel Time)			
***(origin:core/destination:non-core)**	-0.028	-0.026	-0.020
	(0.021)	(0.022)	(0.023)
***(origin:non-core/destination:core)**	-0.101***	-0.111***	-0.127***
	(0.020)	(0.021)	(0.022)
*** Within Region**	-0.023	-0.016	0.007
	(0.020)	(0.020)	(0.020)
Origin-Dest Fixed Effect	✓	✓	✓
Year Fixed Effect	✓	✓	✓
Origin-year Fixed Effect	✓	✓	✓
Destination-year Fixed Effect	✓	✓	✓
Observations	849,660	849,660	849,660
R-squared	0.935	0.919	0.906
N. origin-destination pair	56,644	56,644	56,644

Notes: *** p<0.01, ** p<0.05, * p<0.1. Dependent variable of column (1)-(3) is log of the number of migrants in each district-pair for each demographic category. Core areas are defined as districts in Seoul metropolitan areas, which are districts of Incheon, Seoul, or Gyeonggi provinces. Non-core areas are all other districts. Standard errors are clustered at origin-destination-pair-level.

Table 13: Effect of Travel Time between Origin and Destination Districts on Migration: Couple

	(1)	(2)	(3)
VARIABLES	log of Number of Migrants		
	(Couple)	(with Kid(s))	(without Kids(s))
Panel A.			
log(Travel Time)	-0.019	-0.053***	0.006
	(0.015)	(0.015)	(0.013)
Observations	849660	849660	849660
Adj R2	0.907	0.903	0.852
Panel B.			
log(Travel Time)			
***(origin:core/destination:non-core)**	-0.069***	-0.119***	-0.013
	(0.025)	(0.026)	(0.021)
***(origin:non-core/destination:core)**	-0.031	-0.043*	-0.026
	(0.023)	(0.023)	(0.020)
*** Within Region**	0.039*	-0.002	0.055***
	(0.022)	(0.021)	(0.019)
Origin-Dest Fixed Effect	✓	✓	✓
Year Fixed Effect	✓	✓	✓
Origin-year Fixed Effect	✓	✓	✓
Destination-year Fixed Effect	✓	✓	✓
Observations	849660	849660	849660
Adj R2	0.907	0.903	0.852
N. origin-destination pair	56,644	56,644	56,644

Notes: *** p<0.01, ** p<0.05, * p<0.1. Dependent variable of column (1)-(3) is log of the number of migrants in each district-pair for each demographic category. Core areas are defined as districts in Seoul metropolitan areas, which are districts of Incheon, Seoul, or Gyeonggi provinces. Non-core areas are all other districts. Standard errors are clustered at origin-destination-pair-level.

Table 14: Parameters

Parameter	**Value**	**Source**
Elasticity of Substitution in Production Function	$\sigma = 0.82$	Bae (2014)
Labor Share in Cobb-Douglas Function	$\beta = 0.66$	Bae (2014)
Labor Force Participation Cost CDF Parameter	$\iota = 0.9$	Calibrated

Table 15: Quantitative Decomposition of the Impact of HSR

	Core	**Non-core**
(Reduced Form) Impact on the Gender Employment Gap ($\frac{\partial(\frac{N^m_{jt}}{N^{fe}_{jt}})}{\partial HSR}$)	-19%	-16.2%
(Reduced Form) Impact on the Gender Wage Gap ($\frac{\partial(\frac{W^m_{jt}}{W^{fe}_{jt}})}{\partial HSR}$)	-16%	0%
(Estimated) Impact on Labor Force Participation Cost ($\frac{\partial(1+T_{jt})}{\partial HSR}$)	-4.75%	-16.71%
(Estimated) Impact on Relative Productivity ($\frac{\partial(\frac{\Psi_{Mjt}}{\Psi_{Wjt}})}{\partial HSR}$)	-30.59%	-10.63%

Table 16: Robustness Check: Sub-sample Analysis, without Sejong Special City

	(1)	(2)	(3)
		IV Results	
		log(Employment)	
	(Male)	(Female)	(Sex Ratio)
Treat*Non-core	0.117*	0.207***	-0.164**
	(0.061)	(0.059)	(0.069)
Treat*Core	-0.438***	-0.295***	-0.199***
	(0.070)	(0.090)	(0.055)
Obs	236	236	236

	(4)	(5)	(6)
		IV Results	
		log(Wage)	
	(Male)	(Female)	(Wage Gap)
Treat*Non-core	0.072	0.128	0.003
	(0.257)	(0.175)	(0.096)
Treat*Core	0.046	0.203**	-0.158***
	(0.100)	(0.085)	(0.017)
Obs	178	165	161

	(7)	(8)	(9)
		IV Results	
		log of	
	Population	Employment	Establishment
Treat*Non-core	0.180***	0.151**	0.158***
	(0.044)	(0.054)	(0.048)
Treat*Core	-0.207**	-0.377***	-0.410***
	(0.085)	(0.075)	(0.090)
Obs	237	236	237

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors are in parentheses. The coefficients are the second stage regression results of instrumental variable estimation in Section 6.2. Core areas are defined as districts in Seoul metropolitan areas, which are districts of Incheon, Seoul, or Gyeonggi provinces. Non-core areas are all other districts. Standard errors are clustered at province-level.

Table 17: Robustness Check: Sub-sample Analysis, without Districts in Big Cities

	(1)	(2)	(3)
		IV Results	
		log(Employment)	
	(Male)	(Female)	(Sex Ratio)
Treat*Non-core	0.254***	0.313***	-0.130
	(0.045)	(0.044)	(0.089)
Treat*Core	-0.461***	-0.353***	-0.127***
	(0.000)	(0.000)	(0.000)
Obs	163	163	163

	(4)	(5)	(6)
		IV Results	
		log(Wage)	
	(Male)	(Female)	(Wage Gap)
Treat*Non-core	-0.052	0.090	0.060
	(0.384)	(0.280)	(0.144)
Treat*Core	0.156***	0.284***	-0.127***
	(0.000)	(0.000)	(0.000)
Obs	107	93	90

	(7)	(8)	(9)
		IV Results	
		log of (*IV Results*)	
	Population	Employment	Establishment
Treat*Non-core	0.231***	0.272***	0.229***
	(0.024)	(0.034)	(0.032)
Treat*Core	-0.080***	-0.411***	-0.423***
	(0.000)	(0.000)	(0.000)
Obs	164	163	164

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors are in parentheses. Big Cities includes Seoul, Busan, Daejeon, Daegu, Gwangju, Ulsan, Incheon. The coefficients are the second stage regression results of instrumental variable estimation in Section 6.2. Core areas are defined as districts in Seoul metropolitan areas, which are districts of Incheon, Seoul, or Gyeonggi provinces. Non-core areas are all other districts. Standard errors are clustered at province-level.

Table 18: Robustness Check: Sub-sample Analysis, without Districts close to North Korean Border

	(1)	(2)	(3)
		IV Results	
		log(Employment)	
	(Male)	(Female)	(Sex Ratio)
Treat*Non-core	0.115*	0.204***	-0.159**
	(0.060)	(0.058)	(0.067)
Treat*Core	-0.415***	-0.274***	-0.199***
	(0.059)	(0.075)	(0.053)
Obs	214	214	214

	(4)	(5)	(6)
		IV Results	
		log(Wage)	
	(Male)	(Female)	(Wage Gap)
Treat*Non-core	0.030	0.258**	-0.051
	(0.258)	(0.087)	(0.082)
Treat*Core	0.053	0.219**	-0.166***
	(0.101)	(0.091)	(0.011)
Obs	167	156	153

	(7)	(8)	(9)
		IV Results	
		log of (*IV Results*)	
	Population	Employment	Establishment
Treat*Non-core	0.189***	0.149**	0.162***
	(0.045)	(0.054)	(0.047)
Treat*Core	-0.200**	-0.355***	-0.384***
	(0.091)	(0.062)	(0.083)
Obs	215	214	215

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors are in parentheses. Districts closer to North Korea include all districts in Gangwon Province and districts in Gyeonggi Province (Paju, Pochun, Yeonchun, Gapyung, Dongduchun). The coefficients are the second stage regression results of instrumental variable estimation in Section 6.2. Core areas are defined as districts in Seoul metropolitan areas, which are districts of Incheon, Seoul, or Gyeonggi provinces. Non-core areas are all other districts. Standard errors are clustered at province-level.

A Appendix: Data

A.1 Census on Establishment

Begining in 1994, the Census on Establishments is an annual population survey that provides business information on all 3.3 million enterprises and establishments in South Korea. This administrative dataset includes each establishment's business information, such as its' 5-digit industry code, number of employees, and geographic information. This detailed information makes it possible to construct a panel data set of job creation[19] and job destruction[20] with detailed geographic and establishment information. For the time being, I focus on the number of employees and the number of establishments at the district-level across different industries or establishments with different sizes.

A.2 Population and Housing Census

The Population and Housing Census is an administrative survey that collects information on all 50 million Koreans and foreign residents in South Korea every five years. The survey includes information on demographics (age, gender, education level), socio-economic status (job status), as well as residential location, job location, commuting patterns, and the residential location of the respondent five years prior to the survey.

This study uses 4 waves (2000, 2005, 2010, 2015) of the Census data, which contain a 2% random sample from the population survey. The key variables used in this paper are current residential location, residential location 5 years prior, commute location, commuting mode, and other socio-demographic information such as education level, gender, and marriage.

A.3 Internal Migration Statistics

The Internal Migration statistics is an administrative dataset of residential migration for establishing population policies. The data is collected when a household moves to a zip-code outside of its original zip-code. This makes it possible to track all the across-zipcode migrants in South Korea. The information is collected at the household-level with demographic information (gender and age) on all household members who move together. Importantly, this data contains information on the origin zip-code of each migrant, which lets us track area-to-area migration flows each year.

[19] The number of jobs created each year due to either firms' entrance or expansion of employment of incumbents.

[20] The number of jobs that disappeared each year due to either exit of firms or declines in employment of surviving firms.

By aggregating 88,248,353 observations of annual (2001-2015) household-level data to annual county-level origin-destination (o/d) data, this paper uses of 849,600 [21] observations of o/d migration flows. I focus on the number of migrants, the number of female/male migrants, the number of migrants across different age groups (under 20s, 20s, 30s, 40s, 50s, and over 60s) to see the heterogeneous migration patterns across different demographics.

A.4 Korean Labor and Income Panel Study, KLIPS

KLIPS (Korean Labor & Income Panel Study) is a longitudinal survey of the labor market and income activities of households and individuals residing in urban areas. The 1st Wave of the KLIPS was launched by the KLI (Korea Labor Institute) in 1998, and the survey is conducted annually. Around 5,000 households and 15,000 individuals are surveyed every year.

[21] 849,600 = 238 counties *238 counties *15 years. Note that within county migration is also included in the observation.

B Appendix: Data Construction

B.1 District-level Wage Index

Since the administrative wage data during the sample period is not available, wage index at district-level data is imputed using Korea Labor and Income Panel Study (KLIPS). The individual panel survey provides individuals' rich information about where people live and work, socio-economic characteristics as well as annual labor earnings. To compute the wage index at district-level for before and after HSR construction, I use individual-level wage data, and get the district-fixed effects as the wage index. I make use of four years (2000-2003) for pre-HSR period, and five years (2011-2015) for post-HSR period. Table A1 presents the summary statistics for the original source data.

For an individual i, living in district j, at time $t \in \{before, after\}$, the district-level wage ($\hat{w_{jt}}$) is estimated as follows:

$$w_{ijt} = \alpha + \beta X_{ijt} + \hat{w_{jt}} + \epsilon_{ijt} \tag{17}$$

where X_{ijt} are the vector of individual characteristics such as years of education, age, gender, and experience.

Table A1: Summary Statistics of Wage Data (Source: KLIPS)

Variable	Obs	Mean	Std. Dev.	Min	Max
logwage (pre-HSR)	23,517	4.61	0.61	0.69	7.53
logwage (post-HSR)	28,156	5.29	0.66	1.79	8.29

B.2 A Travel Time Matrix between Cities to Cities

In this study, the reduction in transportation costs induced by the KTX expansion is proxied by the reduction in travel time between cities. I calculate travel time between each county-pair over each year, which consists of a time-varying travel time matrix with an element located in the row j; the column k represents the minimum travel time from a county j to a county k, at a given time t.[22] As of 2010, there are 228 districts, which in turn result in 51,756[23] unidirectional relations between the counties. The travel time within districts is assumed to be zero, as the focus of this paper is on the inter-city movement rather than intra-city movement.

I simplify the travel time calculation by assuming that the distance between each county pair is time fixed, whereas the speed of the transportation modes changes over time as the KTX networks expand. Under this simplifying assumption, calculating the travel time before KTX expansion is relatively easy as the travel time is the linear distance between the centroid of origin and destination county divided by the average speed of an automobile or bus, which is assumed to be 90km/hour, the highway speed limit in South Korea.[24]

The changes in travel time induced by KTX expansion are calculated as follows. First, I construct KTX networks by using the Korean Transportation Database, collected by the Korea Transport Institute (KTDB). The database provides information on the exact location of all train stations, railroad networks, and the opening date of each station. Combining this information with the KTX train schedule available online, I calculate estimated travel time from a station to every other station connected by KTX networks.[25] Finally, the travel time from an origin county to a destination county is calculated as the sum of 1) the estimated time from the centroid of the origin county to departure KTX station; 2) the estimated travel time by KTX from the departure station to the arrival station; and 3) the estimated travel time from the arrival KTX station to the centroid of the Destination county.

Finally, the time-varying minimum travel time matrix is constructed by using the following algorithm. Whenever there was an expansion in KTX networks (e.g., 2004, 2010, 2011, 2012) I compare the travel time before the expansion of KTX to the travel time after the expansion and

[22]The timed travel between the two districts is assumed to be symmetric.

[23](district-district pairs) - (pairs with its own) = (228*228) - 228.

[24]The underlying assumption is that the motor is the closest substitute of the KTX. This is because travel by other transportation modes such as traditional railway or domestic flights takes longer than by car, in the South Korean context (KTI, 2013).

[25]To simplify the calculation, I first calculate the linear distances between each station and divide the distance by the average train speed between each departure-arrival station.

take the smaller value for every county-pair. Before 2004, as KTX had not yet been introduced, the minimum travel time is calculated as the linear distance divided by the speed of an automobile. For 2004, when the KTX was first introduced, the previous travel time is compared to the travel time accounting for the new KTX networks and the smaller travel time is taken between the two. Whenever the KTX network is expanded, the algorithm repeats the same procedure for every county-pair to create the time-varying minimum travel time matrix.

C Appendix: Regression on Migration Flow

Let M_{odt} be the number of migrants from region o to region d in year t. The gravity model for M_{odt} takes the following form:

$$M_{odt} = aPop_{ot}^{\alpha_o}Emp_{ot}^{\beta_o}Pop_{dt}^{\alpha_d}Emp_{dt}^{\beta_d}D_{odt}^{\gamma} \tag{18}$$

where Pop_{ot} (Pop_{dt}) is the population of an origin o (destination d) at time t, Emp_{ot} (Emp_{dt}) is the total employment of origin o (destination d) at time t, and D_{odt} means travel costs (in our context, travel time) between origion (o) and destination (d). The intuition for the model is that the migration flow between an origin and a destination county-pair increases if the size (population, employment) of the origin or the destination county is big and the distance between the two is close enough.

Taking the logarithms of equation 18 we get:

$$log(M_{odt}) = log(a)+\alpha_o log(Pop_{ot})+\alpha_d log(Pop_{dt})+\beta_o log(Emp_{ot})+\beta_d log(Emp_{dt})+\gamma log(D_{odt}) \tag{19}$$

Our coefficient of interest in equation (19) is γ, which can be interpreted as the percentage changes in the number of bilateral migrants associated with a 1% change in travel time.

However, a naive estimation of equation (19) would cause an endogeneity problem. For example, if the placement of HSR specifically was targetting counties which grow or decline, then the causal inference between the travel time reduction and migration flow would be threatened.

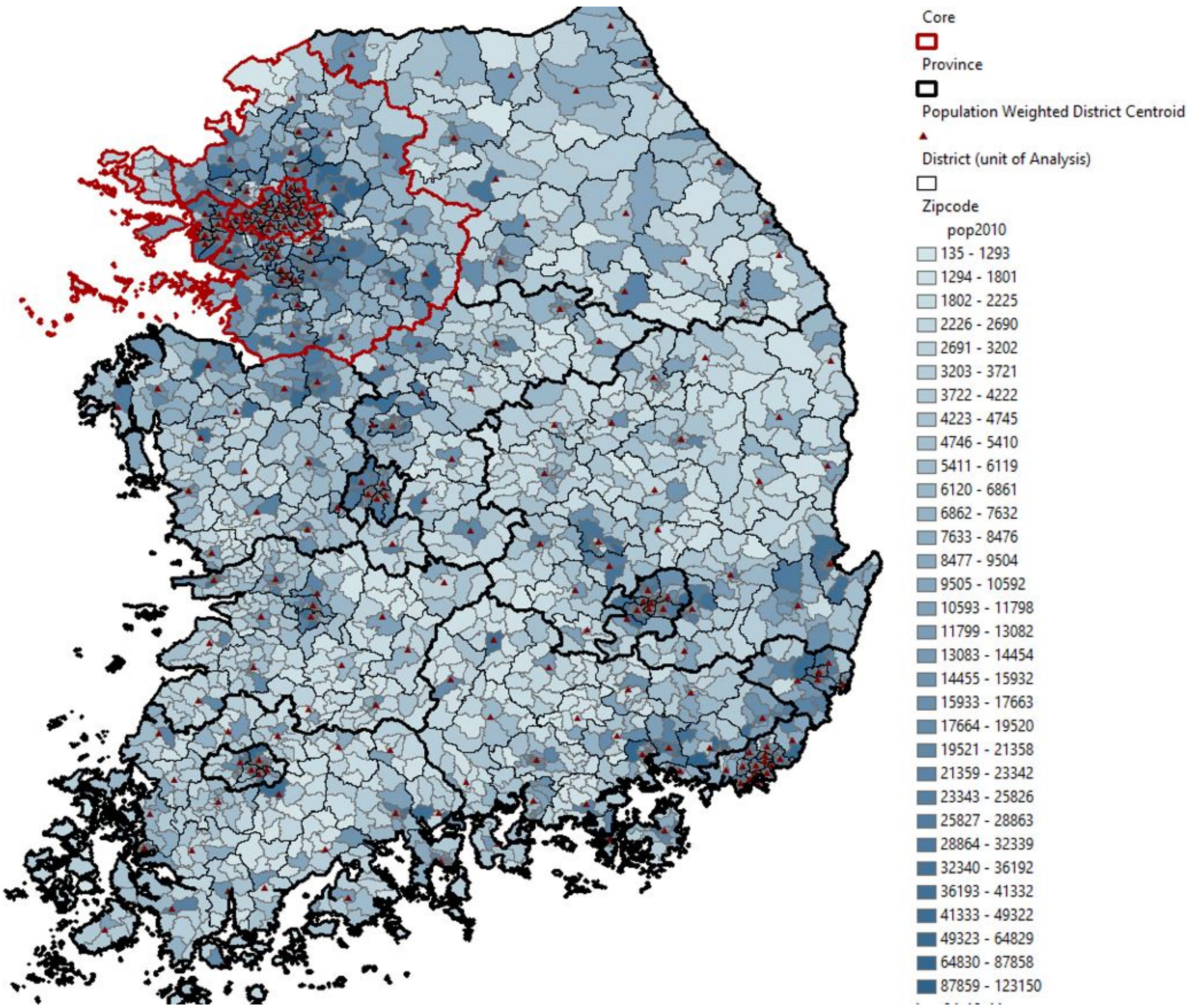
Keeping in mind the potential endogeneity issue, the main regression equation is [26]

$$log(M_{odt}) = \alpha + \gamma log(TravelTime_{odt}) + \eta_{od} + \nu_{ot} + \theta_{dt} + \epsilon_{odt} \tag{20}$$

where, $log(TravelTime_{odt})$ is the travel time between origin and destination counties at time t, η_{od} is the origin-destination pair fixed effect, η_{ot} is the origin-year fixed effect, and θ_{dt} is the destination-year fixed effect.

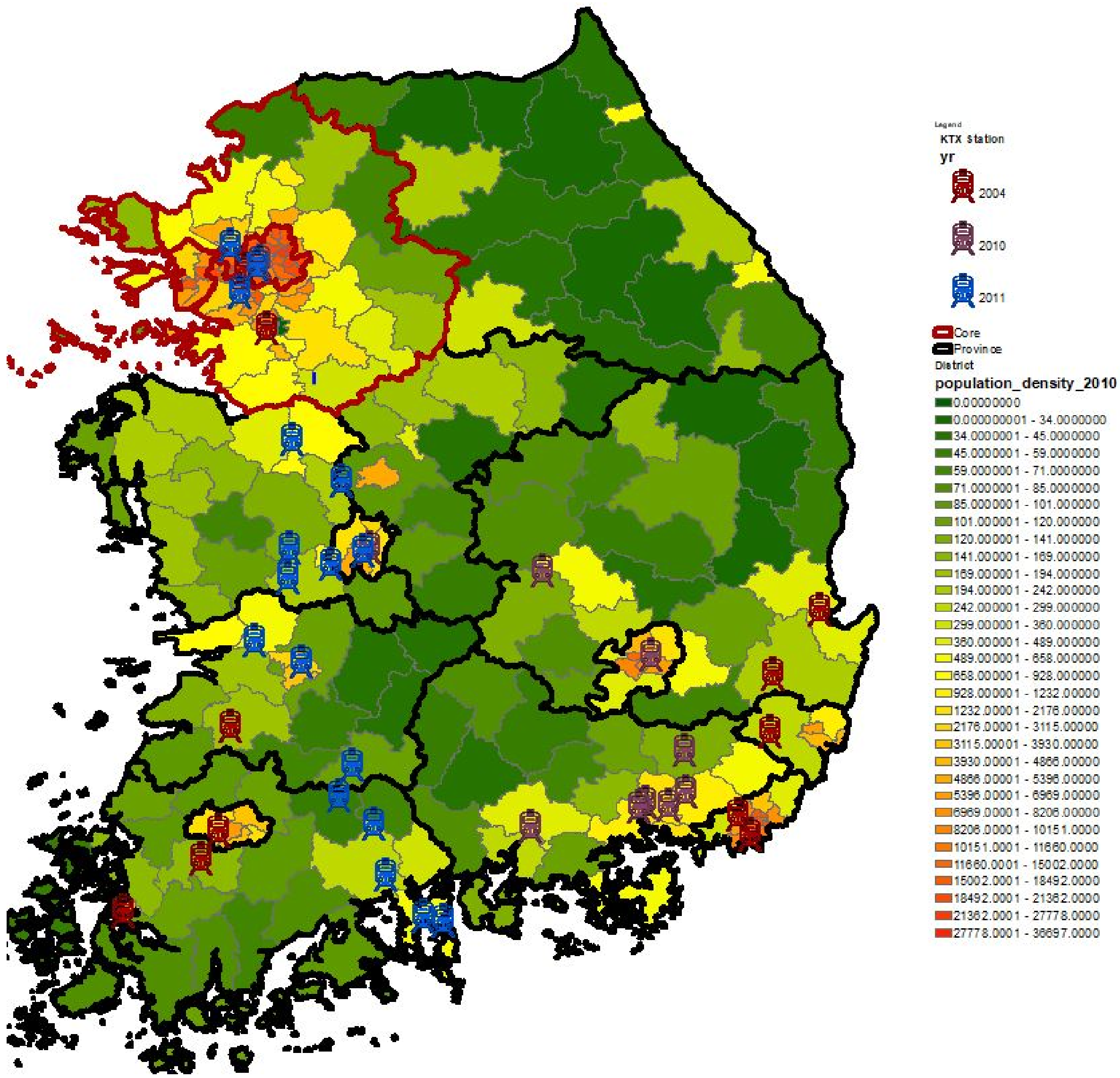
[26] The same empirical specification is also used in Morten and Oliverira (2017)

Figure A1: Population-weighted Centroid



Note: Color of the map represents zip-code-level 2010 population.

Figure A2: KTX Network Expansion



Note: Color represents the population density of districts in 2010. Districts within the thick red lines are defined as core areas; districts in Seoul metropolitan areas, which are districts of Incheon, Seoul, or Gyeonggi provinces.

Table A2: History of the KTX Station Expansion

Date	Line	From/To	Project Type	Station
Apr/2004	Gyeongbu	Seoul/Daegu*	New line	Seoul, Hangshin, Youngdengpo, Gwangmeng Suwon, Cheonan-Asan, Daejeon, Dong-Daegu
		Daegu/Busan**	New line	Gupo, Milyang, Busan
	Honam**	Daejeon/Mokpo	New line	Yongsan, Seo-Daejeon, Gyeryong, Nonsan Iksan, Jeongeup, Gwangju-Songjeong, Naju, Mokpo
Dec/2010	Gyeongbu*	Seoul/Daegu	Add-station	Osong
		Daegu/Busan	Add-station Improve speed	Shin-Gyeongju, Gimcheon(Gumi), Ulsan Busan
	Gyeongjeon**	Milyang/Masan	New line	Jinyoung, Changwon-Joongang, Changwon, Masan
Oct/2011	Jeonla**	Ilsan/Yeosu	New line	Jeonju, Namwon, Guryegu, Sooncheon, Yeocheon Yeosu-expo, Goksung
Dec/2012	Gyeongjeon**	Masan/Jingu	Add-station	Jinju

*High Speed Railroad ($V_{max} = 305km/h$)
**Electrified conventional railway directly connected with HSR ($V_{max} < 180km/h$)
Source: Korail

Table A3: The Inferred impact of HSR on Relative productivity and female labor participation costs

A. Core Areas

A.1 Impact on the Gender Gap in Employment

	Female Labor Participation Cost $(1+T_{jt})$		
Relative Productivity $(\frac{\Psi_{Mjt}}{\Psi_{Wjt}})$	**Decrease**	**No Changes**	**Increase**
Decrease	(-)	(-)	(?)
No Changes	(-)		
Increase	(?)		

A.2 Impact on the Gender Gap in Employment

	Female Labor Participation Cost $(1+T_{jt})$		
Relative Productivity $(\frac{\Psi_{Mjt}}{\Psi_{Wjt}})$	**Decrease**	**No Changes**	**Increase**
Decrease	(?)	(-)	(-)
No Changes			(-)
Increase			(?)

A.3 Inferred Impact of HSR

	Female Labor Participation Cost $(1+T_{jt})$		
Relative Productivity $(\frac{\Psi_{Mjt}}{\Psi_{Wjt}})$	**Decrease**	**No Changes**	**Increase**
Decrease	✓	✓	✓
No Changes			
Increase			

Table A4: The Inferred impact of HSR on Relative productivity and female labor participation costs

B. Noncore Areas

B.1 Impact on the Gender Gap in Employment

	Female Labor Participation Cost $(1+T_{jt})$		
Relative Productivity $(\frac{\Psi_{Mjt}}{\Psi_{Wjt}})$	**Decrease**	**No Changes**	**Increase**
Decrease	(-)	(-)	(?)
No Changes	(-)		
Increase	(?)		

B.2 Impact on the Gender Gap in Wage

	Female Labor Participation Cost $(1+T_{jt})$		
Relative Productivity $(\frac{\Psi_{Mjt}}{\Psi_{Wjt}})$	**Decrease**	**No Changes**	**Increase**
Decrease	(?)		
No Changes		(.)	
Increase			(?)

B.3 Inferred Impact of HSR

	Female Labor Participation Cost $(1+T_{jt})$		
Relative Productivity $(\frac{\Psi_{Mjt}}{\Psi_{Wjt}})$	**Decrease**	**No Changes**	**Increase**
Decrease	✓		
No Changes			
Increase			

Table A5: Mechanism 1.(2). Effect of HSR on Sectoral Sex Ratio (Male-to-Female Ratio)

Panel A. Male Intensive Sector

	(1)	(2)	(3)	(4)
	Sex Ratio			
	(Transportation)	(Construction)	(Public Admin)	(Manufacturing)
Treat				
**Non-Seoul*	0.379	1.576**	1.038**	-0.828**
	(0.595)	(0.704)	(0.368)	(0.340)
**Seoul*	-5.237	4.825**	1.922	-0.621***
	(4.481)	(2.104)	(2.216)	(0.127)
Sex Ratio(2000)	10.09	7.56	4.40	2.36
Observations	237	237	237	237

Panel B. Female Intensive Sector

	(5)	(6)	(7)	(8)
	Sex Ratio			
	(Retail)	(Education)	(Medical Service)	(Restaurant)
Treat				
**Non-Seoul*	0.027	0.422***	-0.053	0.035
	(0.024)	(0.107)	(0.037)	(0.025)
**Seoul*	-0.015	0.991	-0.040	0.002
	(0.027)	(0.913)	(0.093)	(0.014)
Sex Ratio(2000)	0.88	0.76	0.54	0.50
Observations	237	237	236	237

Notes: *** p<0.01, ** p<0.05, * p<0.1. Standard errors are in parentheses. The coefficients are the second stage regression results of instrumental variable estimation in Section 6.2. Dependent variable of column (1)-(8) is sex ratio of each sector, defined as male employment over female employment. Core areas are defined as districts in Seoul metropolitan areas, which are districts of Incheon, Seoul, or Gyeonggi provinces. Non-core areas are all other districts. Standard errors are clustered at province-level.

Table A6: Residential Space, Housing Ownership, and Housing expenditure (Family Survey, 2012)

	(1) Residential Space	(2) Home ownership=1	(3) Housing expenditure (not home-owners)
Omitted Group: Couple, living in non-core			
Core			
Couple	-2.068	-0.148***	36.461***
	(1.339)	(0.014)	(11.981)
Single Men	-42.792***	-0.544***	-33.199*
	(1.571)	(0.026)	(19.667)
Single Women	-38.482***	-0.478***	-37.101*
	(2.539)	(0.032)	(20.670)
Others (single mom/dad)	-22.065***	-0.361***	21.562
	(1.678)	(0.023)	(19.883)
Non-Core			
Single Men	-32.255***	-0.423***	-88.567***
	(2.300)	(0.028)	(15.653)
Single Women	-22.529***	-0.270***	-78.640***
	(2.537)	(0.030)	(17.976)
Others (single mom/dad)	-15.946***	-0.224***	-44.592***
	(1.592)	(0.022)	(12.502)
Observations	7,755	7,755	3,495
R-squared	0.082	0.096	0.020

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A7: Changes in estimations according to the value of ι (%)

	ι	0	0.1	0.3	0.5	0.7	0.9	1
core	Labor force participation cost	-1874.03	-166.39	-43.16	-19.80	-10.12	-4.75	-2.83
core	Relative Productivity	-42.31	-39.42	-35.24	-32.82	-31.39	-30.59	-30.35
non-core	Labor force participation cost	-1611.42	-154.92	-49.23	-29.16	-21.01	-16.71	-15.27
non-core	Relative Productivity	-19.52	-17.70	-14.88	-13.01	-11.66	-10.63	-10.20