The Effects of Cultural Similarity on Foreign Direct Investment and Productivity of Domestic Firms: Identification from Borders of Chinese Dialect Zones

Sen Ma[1]

Department of Economics

University of Illinois at Urbana Champaign

Preliminary Version

This Draft: 11/30/2016

[1] senma2@illinois.edu; Department of Economics, University of Illinois at Urbana-Champaign, 214 David Kinley Hall 1407 W. Gregory Urbana, Illinois 61801, USA

Abstract: Does cultural similarity affect destination choices of foreign direct investment (FDI)? To answer this question, I provide a novel identification strategy which explores discontinuous changes in investment from Hong Kong, Macau and Taiwan (HMT) to Mainland China at borders of Chinese dialect zones. Mainland China can be geographically separated into various dialect zones. People who speak the same Chinese dialect share the same cultural origin. Thus common dialect can be used to measure cultural similarity among Chinese people. I find that if a location in Mainland China is culturally similar to HMT (speaks the same Chinese dialect as HMT), this location attracts more investment from HMT. Causal effect of common dialect on HMT investment is identified from discontinuous increase in investment at geographical borders of common dialect zones. Share of HMT firms in regions right inside the common dialect border is 5 to 7 percentage points higher than regions right outside the border. In addition, I also find that the discontinuous increase in investment from HMT at the borders is larger in industries that receive less FDI entry-restriction policies. Using common dialect and common dialect interacted with entry-restriction policy as exogenous variations at the dialect borders, I find that presence of HMT firms generates negative horizontal spillover effect on productivity of domestic firms from the same region and same industry. A 1 percentage point increase in share of HMT firms decreases total factor productivity of domestic firms by around 1.2 to 1.4 percent. Validity of identification assumptions are further tested in several placebo tests. As a result, I do not find discontinuous changes in share of foreign firms from regions other than HMT at the dialect borders. Neither do I find discontinuous changes in degree of economic development and geographical conditions at the border.

JEL codes: F21, F23,O19, O24

1. Introduction

Does cultural similarity affect economic exchange? In the international trade literature, there is a widely accepted consensus that cultural similarity does affect trade flows among countries (Melitz, 2008; Felbermayr and Toubal, 2010). However, as for another important form of economic exchange: foreign direct investment (FDI), there is not as much empirical evidence. As a pioneering study, Guiso et al. (2009) proposes that cultural similarity does affect location decisions of FDI because sharing similar culture helps managers to build up trust.

In this study, I employ a new identification strategy to establish the causal relationship between cultural similarity and FDI. To establish the identification strategy, I study a special type of FDI: investment from Hong Kong, Macau and Taiwan (abbreviated as HMT) to mainland China. Empirically, I show that if a location in Mainland China is culturally similar to HMT, this location will attract more investment from HMT. To measure cultural similarity, I use whether the location in Mainland China speaks the same Chinese dialect as HMT. China can be geographically separated into various dialect zones. People who speak the same Chinese dialect share the same cultural origin. Thus common dialect can be used to measure cultural similarity among Chinese people. Figure 1 shows the two Chinese dialect zones studied in this research. Regions within the Cantonese dialect zone speak Cantonese which is the same dialect used by Hong Kong and Macau. Similarly, regions within the Min dialect zone speak the same dialect as the majority of Taiwan people. To address the causal effect of common dialect on investment from HMT, I investigate whether investment from HMT to Mainland China increases discontinuously at the geographical borders of common dialect zones.

Using spatial regression discontinuity design and share of HMT firms at postcode level, I find that share of HMT firms in postcodes right inside the common dialect zones are 5 to 7 percentage points higher than share of HMT firms in postcodes right outside the common dialect zones. [2] I interpret this discontinuous increase in HMT investment at the border as causal effect of cultural similarity on FDI under the assumption that factors other than common dialect change continuously at the border.

To strengthen this identification strategy based on spatial discontinuity, I conduct an additional analysis employing industrial and time variations in Chinese FDI entry-restriction policies. I find that if entry of FDI into a specific industry is restricted by government policy in a specific year, the discontinuous increase in investment from HMT at the common dialect border becomes smaller for that industry. In other words, common dialect exerts larger positive impact on attracting FDI in industries with less entry regulation.

Using casual effect identified from exogenous variations at borders of dialect zones, I conduct additional analysis to investigate the spillover effect of FDI on productivity of domestic firms. Receiving FDI from advanced economy is a development strategy undertaken by many developing countries. However, it is empirically unclear how domestic firms are affected by the presence of FDI. The identification of FDI spillover on domestic firms is notoriously hard because the entry decision of foreign multinationals is obviously endogenous (Lu et al., 2015). I aim to improve identification strategy of this problem by addressing productivity variation of domestic firms at the border of dialect zones.

Specifically, I employ the discontinuous increase in HMT investment at common dialect borders and the interaction between common dialect and entry-restriction policy as exogenous variations to identify the causal effect of investment from HMT on productivity of domestic

[2] A Chinese county is of similar size to a U.S. city.

firms from the same location and industry (horizontal spillover effect). Using sample of all domestic firms, I find that investment from HMT generates negative spillover effect on productivity of domestic firms. A 1 percentage point increase in share of HMT firms is transformed into decrease of around 1.2 to 1.4 percent in total factor productivity (TFP) of domestic firms from the same industry and postcode. Identification of this effect requires the assumption that common dialect affects productivity of domestic firms only through attracting more investment from HMT.

To address the identification assumptions, I conduct several placebo tests. Firstly, I show that share of foreign firms from countries other than HMT do not change discontinuously at the borders. Postcodes that speak the same dialect as HMT do not generally attract more FDI. They attract more FDI only from HMT. Second, I show that postcodes that speak the same dialect as HMT are not economically more developed than postcodes that do not speak the same dialect at the border. Third, to address the concern that geographical factors may determine both formation of dialect borders in history and productivity, I show that major geographical variables, such as elevation and slope, don't increase discontinuously at the border.

This study firstly contributes to the literature on the relationship between cultural similarity or linguistic similarity and economic exchange. Various studies have shown that common language or common culture do affect economic activity, especially patterns of international trade (Rauch and Trindade, 2002; Chiswick, 2008; Felbermayr and Toubal, 2010; Egger and Lassmann, 2012; Falck et al., 2012; Sauter, 2012; Melitz and Toubal, 2014; Egger and Lassmann, 2015). Egger and Lassmann (2015) is the study most closely associated with this study in terms of methodology. Using historical language borders in Switzerland, the authors show that patterns of trade change discontinuously at the border. People from French speaking side are more likely to trade with France while people from German speaking side are more likely to trade with Germany. Nonetheless, there is only a small but growing literature that estimates the effect of cultural similarity on FDI (Guiso et al., 2009; Kim et al. 2015). Yet the association between cultural similarity and FDI discovered in previous studies is unlikely to also imply casual effect without a proper identification strategy. Therefore, my study is one of the first studies that estimate the causal effect of linguistic similarity (cultural similarity) on FDI using discontinuous variation at borders of language zones.

This study also contributes to the estimation of spillover effect of FDI on productivity of domestic firms (Aitken B. J. and Harrison A. E. 1999; Javorcik B. S. 2004; Haskel et al. 2007; Gorodnichenko et al. 2014; Lu et al. 2015). From the theoretical perspective, spillover effect can go either way. On one hand, spillover effect can be positive due to knowledge and human capital spillover from foreign firms to domestic firms. On the other hand, when foreign firms increase production within the same industry, productivity of domestic firms may decrease due to competition and market stealing effect. From the empirical perspective, according to a thorough survey of literature by Havranek and Irsova (2011), results vary broadly across methods and countries. Positive spillover effects on suppliers is statistically robust. Yet spillover effects on buyers and firms from the same industry (horizontal spillover effects) are ambiguous. Also, heterogeneity of firms in terms of absorptive capacity, size, productivity and technological level affects sign of spillover effects (Damijan et al, 2013). In the context of China, using panel data methods, Lin et al. (2009) find that HMT investment generates negative horizontal spillover effects. Using changes in Chinese FDI restriction policies as identification strategy, Lu et al. (2015) also find that FDI (not only from HMT) exerts negative horizontal spillover effects on domestic firms. These findings are generally consistent with the findings of my study.

The rest of this article is organized as follows: the second section introduces background knowledge and data used in this analysis; the third section discusses empirical strategies; the fourth section reports empirical results and the last section discusses limitation and future extensions of this project.

2. Background and Data:

Since 1978, when China initiated the open door policy and became more integrated into the world economy, the country witnessed rapid growth in Foreign Direct Investment (FDI). An important feature of FDI in China is that a large proportion of investment come from regions that share similar Chinese culture with mainland China, such as Hong Kong, Macau and Taiwan (HMT). These regions were historically separated from China and became colonies of Britain, Portugal and Japan. Thus these regions established different economic systems and experienced different paths of economic development. At the moment China opened up, these regions, with well-established market institution, were far more developed than mainland China.

Figure 2 shows the share of investment from Hong Kong and Taiwan[3] among all FDI in mainland China from the 1990s. Share of Hong Kong investment was around 30 percent to 40 percent until 2005 and then increased to around 60 percent later on. Share of Taiwan investment was around 10 percent in the 1990s, but was declining over time. Unique link with China is considered as a factor that explains the large share of Hong Kong and Taiwan investment (Zhang, 2005). Cultural links help to reduce transaction cost, build up trust and protect property

[3] Share of Macau investment is very tiny and thus not emphasized in this study.

rights in an underdeveloped market institution. In this study, I conduct a more detailed analysis to closely investigate how cultural similarity affects investment decisions.

Cultural similarity is measured by whether a specific location in mainland China speaks the same Chinese dialect as HMT. Same dialect does not only indicates lower communication cost, but also implies similar cultural origins of the population. I use maps from "Language Atlas of China" to define regions that use specific dialects in mainland China. "Language Atlas of China" are maps published by Chinese Academy of Social Science and shows geographical distribution of major Chinese dialects. The map is constructed by linguists and anthropologists based on their knowledge from field works on the dialect spoken by the majority of the population for each Chinese village. Figure 3 is an illustration of how borders between different dialects are drawn in these maps. Figure 3 is a map of Jiangyin county[4]. From this map, we can see that the county is separated into multiple villages. Linguists and anthropologists have qualitative information on the dialect used by the majority of the population for these villages[5]. Different dialects are visually represented in the figure by different notations. Then the border of dialect zones are draw between villages that speak different dialects. One limitation of the data is that difference in dialect is qualitative in nature. Thus, I can't calculate quantitative changes in the proportion of population that speak a specific dialect at the border, which is usually not form 0 percent to 100 percent. Therefore, the discontinuous change in dialect at the border is "fuzzy" and no data is available to calculate "fuzziness" of the change.

An interesting feature of dialect border is that it doesn't coincide with current county level administrative borders, which usually determine differences in economic policy. Figure 4

[4] This county is not in the analytical sample. I use this example to illustrate how the borders are generally constructed.
[5] I have data on the final borders drawn by linguists and anthropologists, but do not have village level data.

shows an example of the relationship between dialect borders and county borders. The dark-black line, which indicates dialect border, passes through several counties, separating them into regions that speak different dialects. The points in this figure indicate postcodes. My identification strategy will compare postcodes within the same county on two sides of the border. Thus any policy differentiation at the county level will be controlled.

As an additional identification strategy, I explore both industrial and time variations in Chinese FDI regulation policies, which restrict FDI from entering into certain industries. Lu et al. (2016) shows that changes in these FDI regulation policies indeed affect inflow of FDI into regulated industries. The major goal of FDI entry-restriction policy is to protect domestic firms in the same industry from competition from foreign investment. In 1997, the central government of China published the "Catalog for the Guidance of Foreign Investment Industries", which became the government guidelines for regulating inflows of FDI. Specifically the catalog classify products into four categories: (1) FDI was supported, (2) FDI was permitted, (3) FDI was restricted and (4) FDI was prohibited. The catalog of products listed as restricted to FDI underwent significant changes upon China's accession to WTO since 2001. Therefore I can observe variations across industries and time in regulation policy. Using the catalog of goods that receive FDI entry-restrictions, I check at the two-digit industry level whether any goods produced by a specific industry in a specific year is listed on the catalog. If any goods produced by a two-digit industry are listed as either "restricted to FDI" or "forbidden to FDI", I treat that two-digit industry as being restricted in that specific year. Thus I create an industry-year dummy variable "restriction", which indicates whether a specific industry receives FDI entry-restriction policy in a specific year.

Data on HMT firms is calculated from firm level data from Chinese Industrial Census from 1998 to 2006. The census collects data on all industrial firms with sales above 5 million RMB. One limitation of the data is that firms from Hong Kong, Macau and Taiwan are put into one category. Ideally, I should address changes of Hong Kong and Macau firms at the Cantonese border as well as changes of Taiwan firms at the Min dialect border. However, given the data structure, it is impossible to separate Hong Kong and Macau firms from Taiwan firms. To make the identification strategy at the border work, I make an additional assumption that investment from Taiwan are not affected by the Cantonese border while investment from Hong Kong and Macau are not affected by the Min dialect border[6]. Then, I aggregate the analysis of two borders into one framework and address the changes in investment from Hong Kong, Macau and Taiwan at the Cantonese border and Min dialect border. If the assumptions were true, the changes at Cantonese border reflect changes in investment from Hong Kong and Macau while changes at Min dialect border reflect changes in investment from Taiwan. Firm level variables in real value, such as output, value added, employment and real capital are calculated following the framework and deflators provided by Brandt et al. (2012).

Then, the firm level data is aggregated into postcode level[7] and road distance from each postcode to both dialect borders (Cantonese border and Min dialect border) are calculated. I use the smaller distance among the two to define the distance from a specific postcode to dialect

[6] This assumption can't be directly tested. However, its validity is strengthened by placebo tests introduced later in this paper, which shows that investment from other foreign countries don't change across border. Thus I also expect that investment from a region that is not related to the specific dialect shouldn't be affected by the border.
[7] We use the geographical information of current postcode in this version of analysis. Since there could be some changes in geographical locations of some postcodes across years, this may not accurately reflect the geographical location of each postcode in the period of the data. I will address this issue in future versions and try to identify postcodes that experienced location change across years and see whether the results are still robust.

border. For each postcode, I also calculate distance to Hong Kong, if the nearest border is Cantonese border and distance to Taipei if the nearest border is Min dialect border.

3. Empirical Strategy:

3.1. Estimating the effect of common dialect on investment from Hong Kong, Macau and Taiwan (postcode level analysis):

As the first step of estimation, I follow the spatial regression discontinuity design to estimate discontinuous increase in investment from Hong Kong, Macau and Taiwan (HMT) at borders of the two dialect zones.

To construct the dependent variable of this analysis, I aggregate firm level data and calculate share of HMT firms among all firms in postcode i and year t. Employment share of HMT firms is calculated using the following equation:

$$HMT\_share_{it,employment} = \frac{\sum_{f\in\Omega_{it}} HMT_{fit} \times Employment_{fit}}{\sum_{f\in\Omega_{it}} Employment_{fit}} \qquad (1);$$

where $Employment_{fit}$ measures total employment of firm f in postcode i and year t; HMT_{fit} measures HMT equity share of firm f in postcode i and year t; and Ω_{it} is the set of all firms in postcode i and year t. Output share of HMT firms is also calculated using equation (1) as an alternative measure.

The spatial regression discontinuity empirical model is specified as:

$$HMT\_share_{it} = \beta_0 + \beta_1 T_i + f(D_i) + T_i \times f(D_i) + f(DC_i) + c_j + \eta_t + \epsilon_{it} \qquad (2);$$

where i denotes postcode, t denotes year and j denotes county. HMT_share_{it} denotes share of HMT firms in postcode i and year t. T_i is a dummy variable indicating whether postcode i is located inside the common dialect border. D_i is road distance from postcode i to the nearest dialect border. $f(D_i)$ are polynomials of road distance. $f(DC_i)$ are polynomials of distance from postcode i to Hong Kong (if the nearest dialect border is Cantonese)[8] or Taiwan (if the nearest dialect border is Min dialect). c_j are county fixed effects and η_t are year fixed effects. β_1 thus captures discontinuous changes in share of HMT firms at the dialect border.

I estimate the polynomial model specified as equation (2) with a bandwidth of 40 kilometers, because 40 kilometer is approximately the size of one county in terms of road distance. As robustness checks, I also estimate local linear regression model with a more restrictive bandwidth. The model is specified as:

$$HMT\_share_{it} = \beta_0 + \beta_1 T_i + \beta_2 D_i + \beta_3 T_i \times D_i + f(DC_i) + c_j + \eta_t + \epsilon_{it} \quad (3);$$

Optimal bandwidth is chosen by the cross validation method proposed by Imbens and Lemieux (2008).

3.2. Estimating the effect of common dialect and FDI entry-restrictions on investment from HMT (Firm level analysis):

In the next step, I study heterogeneous effects of common dialect on HMT investment across industries to strengthen my identification strategy. Specifically, I evaluate whether the discontinuous increase in share of HMT firms at the dialect borders varies as the industry receives less entry-restrictions. If the positive effect of common dialect on investment from HMT

[8] Given share of Macau firms is very small compare to share of Hong Kong firms and Macau and Hong Kong are close, I only control distance to Hong Kong when evaluating postcodes located near the Cantonese dialect border.

is causal, we should observe larger positive effect in industries that receive less entry-restrictions. The Chinese government constantly restrict entry of FDI into industries that produce certain types of goods. Upon China's WTO accession in 2001, entry-restriction was relieved for some industries (Lu et al., 2015). Therefore I can use both variation across industries and time in FDI entry-restriction to identify the effect of common dialect interacted with entry-restrictions on HMT investment.

Using firm level data, I estimate the following model:

$$HMT_{fikt} = \beta_0 + \beta_1 T_i + \beta_2 R_{kt} \times T_i + \beta_3 R_{kt} + f(D_i) + T_i \times f(D_i) + f(DC_i) + Z_{fikt} + c_j + \eta_t + \delta_k + \epsilon_{it} \qquad (4);$$

where the dependent variable HMT_{fikt} denotes HMT equity share of firm f in postcode i, industry k and year t. R_{kt} is a dummy variable indicating whether industry k in year t receives FDI entry-restriction or not.[9] Z_{fikt} are firm level control variables and δ_k are industry fixed effects. The parameter of interest is β_2 and I expect β_2 to be negative, indicating that the discontinuous increase in HMT firms at the dialect borders is less in industries that receive FDI entry-restriction. In other words, after FDI regulation is relaxed upon China's WTO accession (decrease in R_{kt}), we should observe more discontinuous increase in HMT firms at the border.

3.3. Estimating effect of Common dialect on productivity of domestic firms:

As the third step of my empirical analysis, I use the discontinuous increase in share of HMT firms at the dialect borders interacted with FDI entry-restriction policy as exogenous variations

[9] The Chinese government lists goods that are either "restricted for FDI" or "forbidden to FDI ". If a good that is produced by industry j is listed as either "restricted" or "forbidden" in year t, I code R_{jt} as 1. Goods that are listed as "forbidden" do not change much over time. Goods listed as "restricted" saw some changes upon China's accession to WTO, which is the major source of variation for this estimation.

and an instrumental variable approach to estimate the effect of HMT investment on productivity of domestic firms from the same industry and postcode (horizontal spillover effect). The empirical model is estimated using sample of all domestic firms. A domestic firm is defined as foreign (including HMT) equity share equals to zero[10].

The dependent variable is Total factor productivity of each firm. To calculate TFP of a specific firm, I firstly follow the method proposed by Brandt et al. (2012). TFP is calculated as:

$$\ln\left(TFP_{ft}\right) = (q_{ft} - \overline{q_t}) - \widetilde{S_{ft}}(l_{ft} - \overline{l_t}) - (1 - \widetilde{S_{ft}})(k_{ft} - \overline{k_t}) \qquad (5);$$

where q_{ft}, l_{ft} and k_{ft} are logarithm of value added, labor and capital of firm f in year t. $\overline{q_t}$, $\overline{l_t}$ and $\overline{k_t}$ are industry average logarithm of value added, labor and capital in year t. Labor is weighted by $\widetilde{S_{ft}}$, which denotes wage share in total value added, while capital is weighted by $(1 - \widetilde{S_{ft}})$. $\widetilde{S_{ft}}$ is calculated as $\widetilde{S_{kt}} = (S_{ft} + \overline{S_t})/2$, where S_{ft} is the wage share in total value added of firm f in year t and $\overline{S_t}$ is industry average wage share in year t. In this specification, each firm is compared with a hypothetical average firm in the industry and productivity deviation from the average firm is captured by TFP.

Since $\widetilde{S_{ft}}$ (labor share in production function) is endogenously chosen by the firm, TFP measured using above equation may be biased. I also follow the framework developed by Olley and Pakes (1996) and Ackerberg et al. (2015) to estimate an unbiased measure of $\widetilde{S_{ft}}$ using non-parametric methods. Appendix 1 shows the detailed procedures to get this alternative measure of TFP.

[10] The main empirical results are similar if domestic firm is defined as foreign equity share smaller than 10 percent.

Then, I examine how changes in productivity are caused by increase in share of HMT firms at the dialect borders using an instrumental variable approach. I estimate the following two-stage model:

$$HMT\_share_{ikt} = \gamma_0 + \gamma_1 T_i + \gamma_2 R_{kt} \times T_i + \gamma_3 R_{kt} + f(D_i) + T_i \times f(D_i) + f(DC_i) + Z_{fikt} + c_j + \eta_t + \delta_k + \epsilon_{it}$$

$$\ln(TFP_{fikt}) = \beta_0 + \beta_1 \widehat{HMT\_share}_{ikt} + \beta_3 R_{kt} + f(D_i) + T_i \times f(D_i) + f(DC_i) + Z_{fikt} + c_j + \eta_t + \delta_k + \epsilon_{it} \qquad (6);$$

where in the main model, I regress TFP of firm f from industry k, postcode i and year t on the share of HMT firms in industry k, postcode i and year t. Then share of HMT firms is instrumented by dummy variable T_i, indicating common dialect and $R_{kt} \times T_i$, indicating interaction between common dialect and FDI entry-restriction policy. All other control variables are the same as equation (4). Thus, I treat common dialect and common dialect interacted with FDI entry-restriction policy as exogenous variations at the border and use these variations to identify β_1, which measures the causal effect of HMT investment on productivity of domestic firms from the same postcode and industry .

Identification of this empirical model requires several assumptions. First, factors other than dialect change continuously at the border such that the discontinuous change in share of HMT firm is only caused by dialect. Second, common dialect affects productivity of domestic firms only through increasing investment from HMT. I will conduct four placebo tests to show the validity of these assumptions. First, it is possible that people speaking one type of dialect have pro-business culture such that they generally attract more investment. To address this concern, I replace dependent variable in equation (2) and (4) to the share of foreign firms from countries other than HMT (for example Japan and Korea). I expect to find no significant discontinuity, indicating that regions inside the borders attract more FDI only from HMT. Second, the degree of economic development might be very different across the border. Therefore I conduct additional placebo tests using measures of total scale of economy (for example total industrial output) as dependent variable in equation (2) and expect to show that there is no discontinuity in the degree of economic development across borders. Third, we are concerned that geography might have played an important role in the history of population migration that determined current dialect border. These geographical factors may still affect firm productivity today. To address this concern, I check several geographical variables, namely elevation and slope, and expect to find no discontinuous changes at the border, indicating that one side of the border is not geographically superior to the other side. Finally, different culture may exert direct effect on productivity instead of affecting productivity through attracting more investment from HMT. Thus I conduct another placebo test and check whether productivity of domestic firms change discontinuously at the border in industries with least influence from HMT (lowest share of HMT firms). Similarity in productivity in these industries indicates that productivity of firms from other industries might be similar without investment from HMT.

4. Empirical Results:

4.1. Graphical Analysis:

Before reporting estimation results, it is helpful to firstly have visualized understanding of the discontinuous changes in share of Hong Kong, Macau and Taiwan (HMT) firms at the dialect borders. Figure 5 shows the average of share of HMT firms by distance to dialect borders. I focus my analysis on postcodes that are within 40 kilometers bandwidth of the dialect borders, because 40 kilometer is approximately the size of one county in terms of road distance. In this figure, the horizontal axis indicates distance to borders with negative value indicating inside the dialect border (common dialect with HMT). Share of HMT firms is standardized by subtracting county mean and divided by county standard error to remove county fixed effects. Figure 5.1 shows employment share, Figure 5.2 shows output share respectively.

From Figure 5.1, we can identify a significant drop in the share of HMT firms at the border when moving from inside the border to outside. We can observe an increasing trend of share of HMT firms within the border (left hand side of the border), probably because Hong Kong is located close to the border from inside. However, this trend did not continue across dialect borders. Share of HMT firms drop discontinuously when we pass through the dialect border to regions that speak different dialect from Hong Kong and Taiwan. Figure 5.2 shows similar empirical patterns when I use output share as an alternative measure of share of HMT firms.

Figure 6 shows average of share of HMT firms by distance to dialect borders with fitted model estimated by the 4^{th} order of polynomials of distance to dialect border. Similar to Figure 5, we observe discontinuous drop in the share of HMT firms at the border of common dialect zones.

As placebo tests, Figure 7 investigates whether we can observe similar discontinuous changes in share of foreign firms from regions other than HMT at the dialect borders. Using the same set up as Figure 5, Figure 7 shows no discontinuous changes in share of firms from other foreign countries at the dialect border, indicating that common dialect increases FDI only from attracting more investment from HMT.

4.2. Main results:

Table 1 reports estimation results of equation (2) and (3) at postcode level. I only report the coefficients on the dummy variable indicating common dialect (β_1) in the table, which is the main treatment effect I try to identify with the spatial regression discontinuity design. I use employment share as the dependent variable in Panel 1 and output share as the dependent variable in Panel 2 respectively. Column (1) to Column (4) are estimated using equation (2). Each column represents a separate estimation with corresponding orders of polynomials of distance to borders as control variables. Column (5) is estimated using equation (3). The optimal bandwidth chosen by cross validation are 30 kilometers for employment share, and 29 kilometers for output share. I also include county fixed effects, year fixed effects, polynomials of distance to Hong Kong (or Taipei) as control variables in all specifications

Panel 1 of Table 1 shows that common dialect discontinuously increases employment share of HMT firms by around 5 to 7 percentage points at the borders of dialect zones. The positive effects are similar in magnitude and statistically significant across all different specifications. Panel 2 shows that output share of HMT firms increase by around 4 to 7 percentage points at the border. All specifications are similar in magnitude and only the model with the 4^{th} order of polynomial as control is statistically insignificant.

Next, I show in Table 2 estimation results of equation (4). The models are estimated at the firm level and the dependent variable is HMT equity share of a specific firm. I report the coefficients on the dummy variable indicating common dialect (β_1) as well as the coefficients on the interaction between common dialect and the dummy variable indicating whether the industry receives entry-restriction policy (β_2). Similar to Table 1, Column (1) through Column (4) represents separate estimations with corresponding orders of polynomials of distance to borders as controls. Column (5) is estimated with local linear model where bandwidth is chosen to be 30 kilometers.

From Table 2, we can conclude that the coefficients on the common dialect dummy are generally positive and statistically significant indicating that common dialect discontinuously increases HMT equity share at dialect borders in industries that do not receive entry-restriction policies. The coefficients on the interaction between the common dialect dummy and entry-restriction policy dummy are negative and statistically significant, indicating that equity share of HMT firms do not increase as much at the border if the industry receives entry-restriction policy. The findings are consistent with the hypothesis that common dialect generates positive casual effect on HMT investment.

Then, I use the discontinuous increase in share of HMT firms at the dialect borders and interaction between common dialect and FDI entry-restriction policy as exogenous variations to estimate the spillover effect of investment from HMT on productivity of domestic firms. To be specific, I estimate equation (6) by regressing TFP on share of HMT firms in the same industry (2 digit) with share of HMT firms instrumented by common dialect and the interaction between common dialect and entry-restriction policy. All models are estimated using local linear model with bandwidth chosen to be 30 kilometers. Panel 1 and Panel 2 of Table 3 show estimation results with 2SLS. I use employment share in Panel 1and output share in Panel 2 as the dependent variables in the first stage respectively. For both panels, I report the estimated results of β_1, which measures the spillover effect of investment from HMT on TFP of domestic firms, as well as the coefficient on common dialect (γ_1) and the interaction between common dialect and entry-restriction policy (γ_2) in the first stage of 2SLS estimation. In Column (1), I use TFP estimated following Brandt et al. (2012) and in Column (2) TFP estimated non-parametrically following Ackerberg et al. (2015) as the dependent variables. All models also include firms' years in business, logarithm of total output, labor capital ratio, a dummy variable indicating whether the firm involves in export and a dummy variable for state owned firms as controls.

Column 1 of Table 3 shows that if share of HMT firms (employment or output share) increase by 1 percentage point, TFP of domestic firms in the same industry decrease by around 2.1 to 2.6 percent. The effect is statistically significant indicating negative spillover effect of investment from HMT on productivity of domestic firms. In the first stage of 2SLS, the coefficient on common dialect is positive and statistically significant and the coefficient on the interaction term is negative and statistically significant, indicating that the instruments are strong and identification is valid. In Column 2, where an alternative measure of TFP is used, I find qualitatively similar results, even though the magnitude of spillover effect is smaller. If share of HMT firms increases by 1 percentage point, TFP of domestic firms in the same industry decrease by around 1.1 to 1.5 percent.

4.3 Placebo Tests:

In order to show the validity of identification assumptions specified in section 3, I conduct several placebo tests and the results are shown in Table 4 through Table 9.

Table 4 shows estimation results of equation (2) and (3) with share of other foreign firms as the dependent variables. Model specifications are the same as Table 1. As expected, I do not find significant discontinuity at the dialect border for most of the models. Only the 3rd order polynomial specification has significant effect but the direction is negative. Thus I can claim that regions inside the dialect border do not generally attract more FDI. They attract more FDI only from HMT.

Table 5 shows estimation results of equation (4) with other foreign equity share as the dependent variables. Model specifications are the same as Table 2. Neither do I find consistent significant positive coefficients on common dialect nor negative coefficients on the interaction between common dialect and entry-restriction policy, indicating that the dialect border do not discontinuously affect investment decisions of foreign firms from regions other than HMT.

Table 6 shows estimation results of equation (2) and (3) with measures of the size of total industrial sector as the dependent variables. I can find discontinuous changes in many model specifications. Yet the direction of changes are not consistent. Thus there is no evidence to suggest that regions inside the dialect border are more economically developed than regions outside the border and thus they attract more investment.

Table 7 shows estimation results of equation (2) with elevation and slope as the dependent variables. Since these geographical variables don't change across time, I estimate a cross sectional version of equation (2). Panel 1 shows that there is no significant discontinuous change in elevation across border. Panel 2 shows that there are some discontinuous changes in slope across border, but the direction of changes are not consistent. Therefore there is no evidence to suggest that regions in one side of the border are geographically superior to the other side. Thus it is unlikely that better geography determines both productivity and historical dialect borders.

Finally, Table 8 shows estimation results of equation (3) with TFP of domestic firms as the dependent variable. This exercise can be regarded as estimating a "reduced form" model to examine the direct effect of common dialect border on productivity. To investigate the hypothetical question whether productivity of domestic firms differ at the dialect borders without investment from HMT, I estimate this model using sample of firms from industries that receive the least influence from HMT (share of HMT firms smaller than 4%). I report the coefficients on the common dialect dummy variable in the table. I find very small and statistically insignificant changes in productivity of these firms at the dialect border, suggesting that firms might not be different in productivity without investment from HMT.

5. Discussions and Further Extensions:

In this study, I show that foreign investment from Hong Kong, Macau and Taiwan (HMT) generates negative horizontal spillover effect on productivity of domestic firms from the same geographical location. The causal effect is identified through exploring discontinuous changes in HMT investment at the borders of Chinese dialect zones. Detailed mechanisms that drive this negative spillover effect can be further explored using the same empirical framework and data. For example, is negative spillover effect realized through making domestic firms invest less in R&D? Whether the negative spillover effect is affected by ownership type of domestic firms? Whether private firms lose more from the presence of HMT firm in the same industry?

The magnitude of spillover effect estimated by this research can't be directly compared with the literature, because the identification strategy of this study only allows estimation of local spillover effects, i.e. spillover effects on other firms from the same postcode. However, in the literature, spillover effect is usually not restricted to firms from the same location. Also, as shown by Halpern and Muraközy (2007), distance matters when calculating magnitude of spillover effects. Therefore, I expect to find spillover effect larger in magnitude given the fact that firms from the same location are more closely interrelated. In this study, it is estimated that if share of HMT firms increases by 1 percentage point, TFP of domestic firms decreases by around 1.1 to 1.5 percent (using the alternative measure of TFP). In the most similar study where spillover effect is not restricted to the same geographical location (Lin et al, 2009), the negative effect on TFP of domestic firms is estimated to be around 0.7 to 0.9 percent.

Another limitation of this study is that the identification strategy only allows the estimation of spillover effects of HMT firms. However, the spillover effects of HMT firms maybe systematically different from those of other foreign firms as shown by Lin et al. (2009) and Du et al. (2012). Both studies find that non-HMT foreign firms (primarily from OECD countries) generate positive spillover effects, yet HMT firms generate negative (close to zero) spillover effects in China. Thus the findings of this study may not be directly applied to FDI from other countries.

In this study, I mainly focus on the analysis of horizontal spillover effect of HMT investment, i.e. how presence of HMT firms in the same industry affect the performance of domestic firms. However, as shown by Melitz and Toubal (2014), spillover effects are more likely to happen across industries, especially on suppliers of foreign firms. However, spillover effects across industries are unlikely to be identified from the empirical setting of this study, because I rely on regional variation (across the dialect border) in HMT investment to identify spillover effect. It is unlikely that suppliers and buyers of a specific firm is restricted to the area where the firm is located. Also, in my sample, many postcodes specialize in certain industries. Therefore it is not possible to observe spillover effects on other industries in the same geographical location. Thus, unable to identify spillover effects across industries will be one of the major limitations of this study.

Literature:

Ackerberg D. A., Caves K. and Frazer G., 2015, “Identification Properties of Recent Production Function Estimators”, *Econometrica*, forthcoming.

Aitken B. J. and Harrison A. E., 1999, “Do domestic firms benefit from direct foreign investment? Evidence from Venezuela”, *The American Economic Review*, 89(3), 605-618.

Brandt L., Biesebroeck, J.V. and Zhang Y., 2012, “Creative accounting or creative destruction? Firm-level productivity growth in Chinese manufacturing”, *Journal of Development Economics*, 97, 339-351.

Chiswick, B., 2009, “The economics of language for immigrants: An introduction and overview.” *The education of language minority immigrants in the United States*, pp.72-91.

Damijan, J.P., Rojec, M., Majcen, B. and Knell, M., 2013. “Impact of firm heterogeneity on direct and spillover effects of FDI: Micro-evidence from ten transition countries.” *Journal of comparative economics*, 41(3), pp.895-922.

Du, L., Harrison, A. and Jefferson, G.H., 2012. “Testing for horizontal and vertical foreign investment spillovers in China, 1998–2007.” *Journal of Asian Economics*, 23(3), pp.234-243.

Egger, P.H. and Lassmann, A., 2012, “The language effect in international trade: A meta-analysis.” *Economics Letters*, 116(2), pp.221-224.

Egger P. H. and Lassmann A., 2015, “The causal impact of common native language on international trade: Evidence from a spatial regression discontinuity design”, *The Economic Journal*, 125, 699-745.

Falck O., Heblich S., Lameli, A. and Sudekum J., 2012, “Dialects, cultural identity and economic exchange”, *Journal of Urban Economics*, 72, 225-239.

Felbermayr G. J. and Toubal F., 2010, “Cultural proximity and trade”, *European Economic Review*, 54, 279-293.

Gorodnichenko, Y., Svejnar, J. and Terrell, K., 2014, “When does FDI have positive spillovers? Evidence from 17 emerging market economies”, *Journal of comparative economics*, 42, 954-969.

Guiso, L, Paola Sapienza, and Luigi Zingales, 2009, “Cultural biases in economic exchange”, *The Quarterly Journal of Economics*, 124(3), 1095-1131.

Haskel J. E., Pereira S. C. and Slaughter M. J., 2007, “Does inward foreign direct investment boost the productivity of domestic firms?”, *The Review of Economics and Statistics*, 89(3), 482-496.

Havranek, Tomas, and Zuzana Irsova, 2011 "Estimating vertical spillovers from FDI: Why results vary and what the true effect is." *Journal of International Economics*, 85(2): 234-244.

Imbens G. W. and Lemieux T., 2008, “Regression discontinuity designs: A guide to practice”, *Journal of Econometrics,* 142, 615-635.

Javorcik B. S., 2004, “Does foreign direct investment increase the productivity of domestic firms? In search of spillovers through backward linkages”, *The American Economic Review*, 94, 3, 605-627.

Kim M., Liu A., Tuxhorn K., Brown D. and Leblang D., 2015, “Lingua Mercatoria: Language and foreign direct investment”, *International Studies Quarterly*, 59, 330-343.

Halpern, L. and Muraközy, B., 2007. “Does distance matter in spillover? 1.Economics of Transition”, 15(4), pp.781-805.

Lu Y., Tao Z, and Zhu L., 2015, “Identifying FDI Spillovers”, *Working Paper*.

Melitz, Jacques, 2008, "Language and foreign trade." *European Economic Review*, 52(4), 667-699.

Melitz, Jacques, and Farid Toubal, 2014, "Native language, spoken language, translation and trade." Journal of International Economics, 93(2): 351-363.

Olley S. and Pakes A., 1996, “The Dynamics of Productivity in the Telecommunications Equipment Industry”, *Econometrica*, 64, 6, 1263-1297.

Rauch, J.E. and Trindade, V., 2002, “Ethnic Chinese networks in international trade.” *Review of Economics and Statistics*, 84(1), pp.116-130.

Sauter, N., 2012. “Talking trade: language barriers in intra-Canadian commerce.” Empirical Economics, 42(1), pp.301-323.

Wurm, S. A. (Eds.) (1988, c1987) Language atlas of China, Hong Kong : Longman Group print.

Zhang K.H., 2005, “Why does so much FDI for Hong Kong and Taiwan go to Mainland China”, *China Economic Review,* 16, 293-307.

Table 1: Common Dialect on Share of HMT Firms

	(1)	(2)	(3)	(4)	(5)
	1st order	2nd order	3rd order	4th order	Local linear
Panel 1:					
Employment share	0.063**	0.075**	0.048*	0.067*	0.070***
	(0.011)	(0.017)	(0.026)	(0.036)	(0.013)
Optimal Bandwidth					30
Observations					7067
Panel 2:					
Output share	0.046***	0.056***	0.065**	0.044	0.055***
	(0.012)	(0.018)	(0.027)	(0.037)	(0.013)
Optimal Bandwidth					29
Observations					6842
Controls	County fixed effects, year fixed effects, distance to Hong Kong or Taipei				
Observations	9456	9456	9456	9456	

Notes: This table shows the effect of common dialect on share of HMT firms. Column (1) to (4) are estimated using equation (2), Column (5) is estimated using equation (3) with optimal bandwidth shown in the table. Optimal Bandwidth is chosen using cross validation and unit is kilometer. From column (1) to (4), each column represents a separate estimation with corresponding orders of polynomials of distance to border as controls. Estimated coefficients β_1 , which are the measure of discontinuous changes at the border, are reported in the table. Panel 1 uses employment share as the dependent variable. Panel 2 uses output share as the dependent variable respectively. Robust standard errors are in parenthesis. *, ** and *** indicate statistical significance at 10%, 5% and 1% level respectively.

Table 2: Common dialect and FDI entry-restriction on HMT equity share

	(1)	(2)	(3)	(4)	(5)
	1st order	2nd order	3rd order	4th order	Local linear
Panel 1: HMT equity share					
Common Dialect (β_1)	0.038***	0.074***	0.053*	0.042	0.073***
	(0.014)	(0.021)	(0.030)	(0.036)	(0.016)
Common Dialect*FDI restriction (β_2)	-0.041***	-0.043***	-0.043***	-0.044***	-0.034***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.013)
Optimal Bandwidth					30
Observations	77531	77531	77531	77531	58054
Controls	County fixed effects, year fixed effects, industry fixed effects, distance to Hong Kong or Taipei, Firms' years in business, output, capital-labor ratio, whether exporter or state firm.				

Notes: This table shows the effect of common dialect and FDI entry-restriction policy on equity share of HMT firms. Column (1) to (4) are estimated using equation (4). Each column represents a separate estimation with corresponding orders of polynomials of distance to border as controls. Coefficients β_1 and β_2 from equation (4) are shown in the table. Column (5) estimates local linear version of equation (4). Standard errors clustered at firm level are in parenthesis. *, ** and *** indicate statistical significance at 10%, 5% and 1% level respectively.

Table 3: Share of HMT firms on productivity of domestic firms (Two stage IV estimation)

	(1)	(2)
	Ln(TFP)	Ln(TFP) Non-parametric $\beta_L = 0.2793$ $\beta_K = 0.4429$
Panel1:		
Employment Share of HMT firms (β_1)	-2.16***	-1.17**
	(0.78)	(0.48)
First Stage (Common dialect) (γ_1)	0.088***	0.088***
	(0.01)	(0.01)
First Stage (Common dialect*FDI Restriction) (γ_2)	-0.065***	-0.065***
	(0.007)	(0.007)
Panel2:		
Output Share of HMT firms(β_1)	-2.61***	-1.41**
	(0.94)	(0.58)
First Stage (Common dialect) (γ_1)	0.065***	0.065***
	(0.010)	(0.010)
First Stage (Common dialect*FDI Restriction) (γ_2)	-0.055***	-0.055***
	(0.007)	(0.007)
Observations	33589	33589
Controls	County fixed effects, year fixed effects, industry fixed effects, distance to Hong Kong or Taipei, Firms' years in business, output, capital-labor ratio, whether exporter or state firm.	

Notes: This table shows the effect of HMT investment on productivity of domestic firms from the same postcode and industry. The model is estimated using local linear version of equation (6) with bandwidth chosen to be 30 kilometers. Panel 1 uses employment share as the endogenous independent variable while panel 2 uses output share as the endogenous independent variable. Column (1) uses TFP calculated following Brandt et al. (2012) as the dependent variable. Column (2) uses TFP calculated using non-parametric method following Ackerberg et al. (2015) as the dependent variable. Coefficients reported in the models are β_1 (coefficient on the endogenous variable in second stage), γ_1 and γ_2 (coefficients on instruments in the first stage). Standard errors clustered at firm level are in parenthesis. *, ** and *** indicate statistical significance at 10%, 5% and 1% level respectively.

Table 4: Common Dialect on Share of Other Foreign Firms (Placebo 1.1)

	(1)	(2)	(3)	(4)	(5)
	1st order	2nd order	3rd order	4th order	Local linear
Panel 1:					
Employment share	0.001	-0.01	-0.037**	0.015	-0.002
	(0.01)	(0.01)	(0.016)	(0.020)	(0.018)
Optimal Bandwidth					11
/observations					2304
Panel 2:					
Output share	0.0004	0.0006	-0.078***	-0.020	-0.025
	(0.01)	(0.013)	(0.018)	(0.022)	(0.023)
Optimal Bandwidth					11
/observations					2304
Controls	County fixed effects, year fixed effects, distance to Hong Kong or Taipei				
Observations	9456	9456	9456	9456	

Notes: This table shows effect of Common Dialect on share of other foreign firms. Column (1) to (4) are estimated using equation (2), Column (5) is estimated using equation (3) with optimal bandwidth shown in the table. Optimal Bandwidth are chosen using cross validation and unit is kilometer. From column (1) to (4), each column represents a separate estimation with corresponding orders of polynomials of distance to border as controls. Estimated coefficients β_1 , which are the measure of discontinuous changes at the border, are reported in the table. Panel 1 uses employment share as the dependent variable. Panel 2 uses output as the dependent variable. Robust standard errors are in parenthesis. *, ** and *** indicate statistical significance at 10%, 5% and 1% level respectively.

Table 5: Common dialect and FDI entry-restriction on Other foreign firms' equity share (Placebo 1.2)

	(1)	(2)	(3)	(4)	(5)
	1st order	2nd order	3rd order	4th order	Local linear
Panel 1: Other Foreign firms' equity share					
Common Dialect (β_1)	0.010	0.027**	0.003	-0.034	0.018*
	(0.009)	(0.014)	(0.02)	(0.023)	(0.01)
Common Dialect*FDI restriction (β_2)	-0.008	-0.008	-0.009	-0.011	-0.007
	(0.007)	(0.007)	(0.007)	(0.007)	(0.008)
Optimal Bandwidth					30
Observations	77531	77531	77531	77531	58054
Controls	County fixed effects, year fixed effects, industry fixed effects, distance to Hong Kong or Taipei, Firms' years in business, output, capital-labor ratio, whether exporter or state firm.				

Notes: This table shows the effect of common dialect and FDI entry-restriction policy on equity share of other foreign firms. Column (1) to (4) are estimated using equation (4). Each column represents a separate estimation with corresponding orders of polynomials of distance to border as controls. Coefficients β_1 and β_2 from equation (4) are shown in the table. Column (5) estimates local linear version of equation (4). Standard errors clustered at firm level are in parenthesis. *, ** and *** indicate statistical significance at 10%, 5% and 1% level respectively.

Table 6: Common Dialect on Total Industrial Production (Placebo 2)

	(1)	(2)	(3)	(4)	(5)
	1^{st} order	2^{nd} order	3^{rd} order	4th order	Local linear
Panel 1:					
Employment	11.76	68.37	-1165.75*	-662.86	-1415.75**
	(332.1)	(449.1)	(608.8)	(831.97)	(688.14)
Optimal Bandwith					10
/observations					2111
Panel 2:					
Output	2.38**	2.53*	-3.34	-3.75	-3.98**
	(1.12)	(1.40)	(2.07)	(2.33)	(1.75)
Optimal Bandwith					9
/observations					1840
Controls	County fixed effects, year fixed effects, distance to Hong Kong or Taipei				
Observations	9456	9456	9456	9456	

Notes: This table shows effect of common dialect on total industrial production of postcode i. Column (1) to (4) are estimated using equation (2), Column (5) is estimated using equation (3) with optimal bandwidth shown in the table. Optimal Bandwidth is chosen using cross validation and unit is kilometer. From column (1) to (4), each column represents a separate estimation with corresponding orders of polynomials of distance to border as controls. Estimated coefficients β_1 , which are the measure of discontinuous changes at the border, are reported in the table. Panel 1 uses total employment as the dependent variable. Panel 2 and Panel 3 use total output and total capital as the dependent variable respectively. *, ** and *** indicate statistical significance at 10%, 5% and 1% level respectively.

Table 7: Common dialect on Geographical Attributes (Placebo 3)

	(1)	(2)	(3)	(4)
	1st order	2nd order	3rd order	4th order
Panel 1:				
Elevation	13.32	16.11	9.86	22.81
	(9.58)	(12.42)	(17.33)	(22.45)
Panel 2:				
Slope	0.56***	0.42*	-0.12	-0.72*
	(0.18)	(0.25)	(0.32)	(0.43)
Controls	County fixed effects, year fixed effects, distance to Hong Kong or Taipei			
Observations	1475	1475	1475	1475

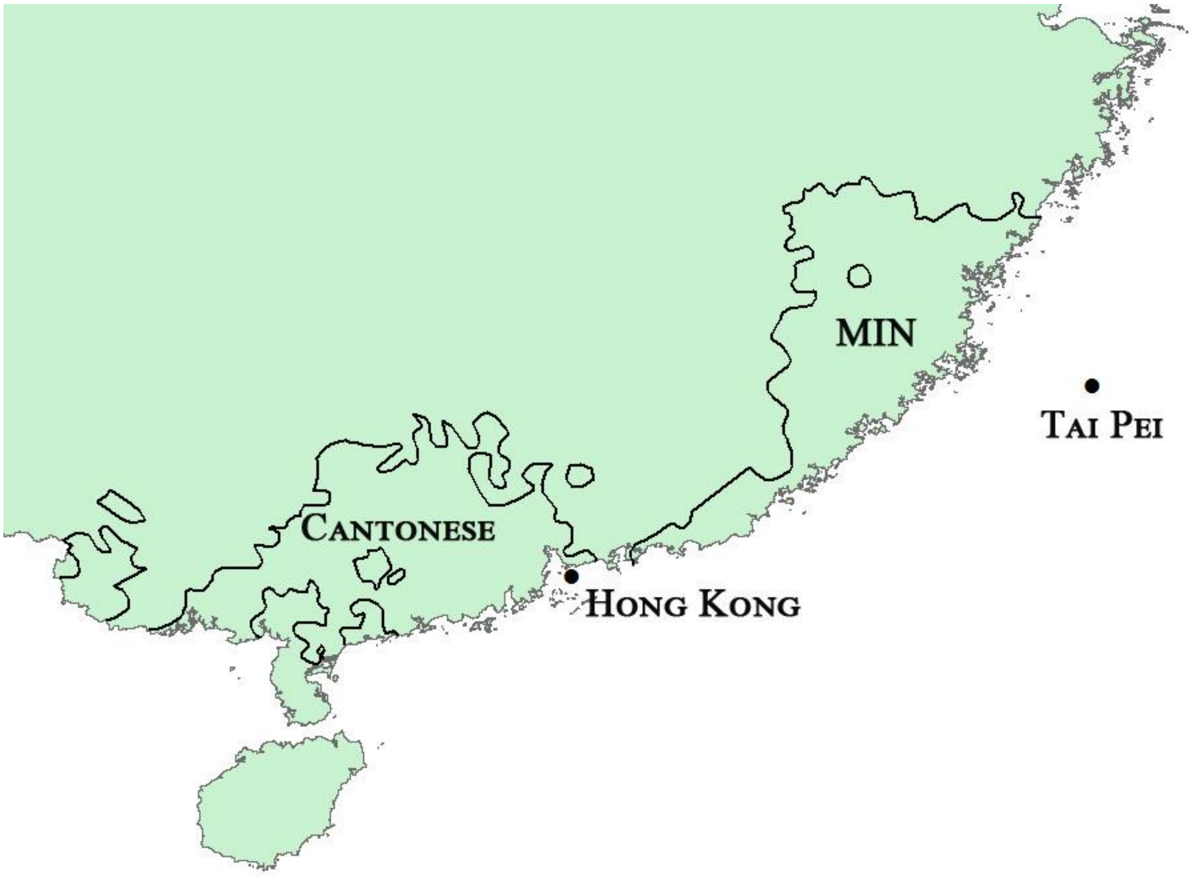
Notes: This table shows effect of Common Dialect on elevation and slope of postcode i. Column (1) to (4) are estimated using equation (2). From column (1) to (4), each column represents a separate estimation with corresponding orders of polynomials of distance to border as controls. Estimated coefficients β_1 , which are the measure of discontinuous changes at the border, are reported in the table. Panel 1 uses elevation as the dependent variable. Panel 2 uses slope as the dependent variable. *, ** and *** indicate statistical significance at 10%, 5% and 1% level respectively.

Table 8: Common Dialect on Productivity (Firms from industries with low HMT share) (Placebo 4)

	Dependent Variables	
	(1)	(2)
	Ln(TFP)	Ln(TFP) Non-parametric $\beta_L = 0.2793$ $\beta_K = 0.4429$
Panel1 (Reduced form)		
Common dialect	-0.002	-0.039
	(0.11)	(0.075)
Observations	2806	2806

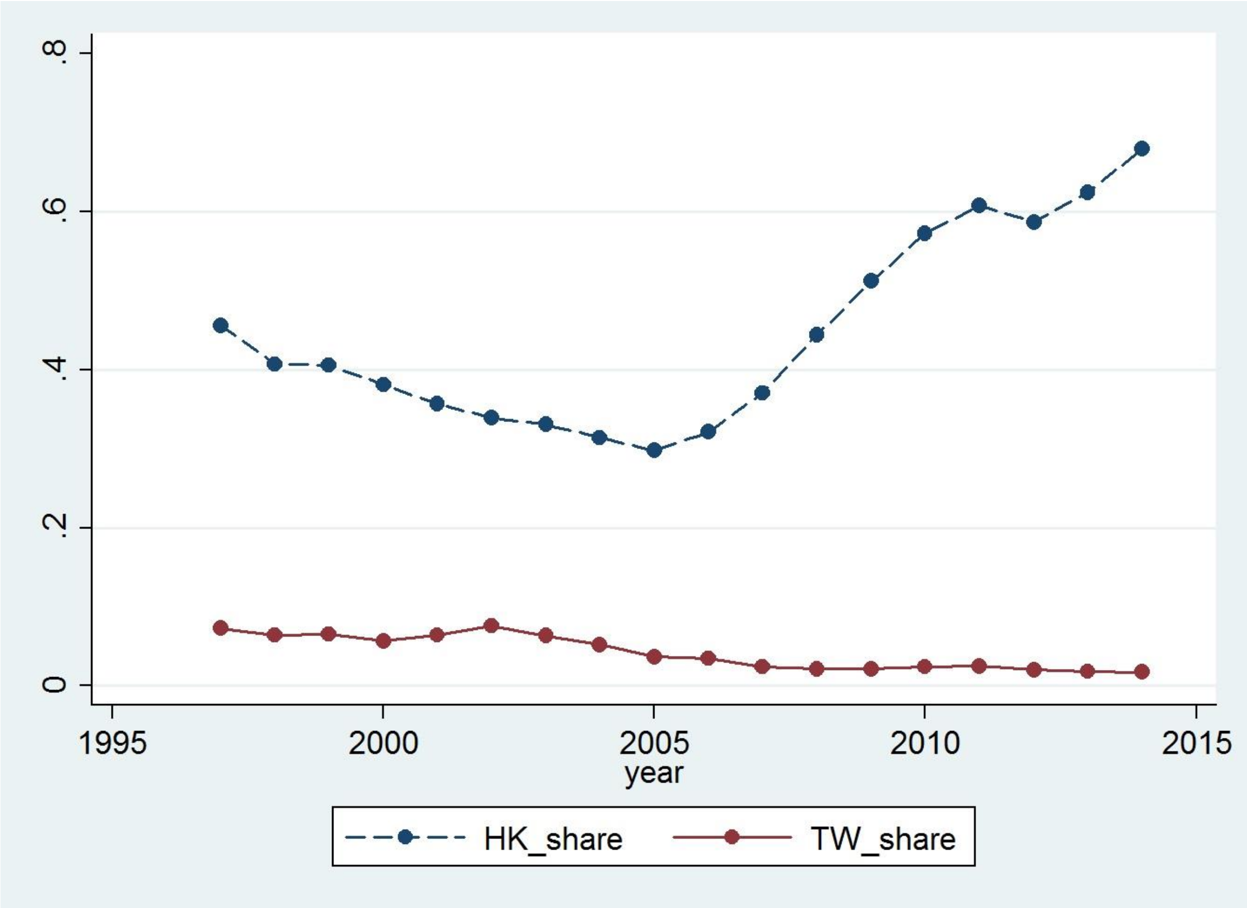
Notes: This table shows effect of common dialect on productivity of domestic firm from industries with low influence of HMT firms (share of HMT firms<4%). The model is estimated using equation (3) and coefficients on the dummy variable indicating common dialect is reported in the table (β_1). Robust standard errors in parenthesis. *, ** and *** indicate statistical significance at 10%, 5% and 1% level respectively.

Figure 1: Cantonese Dialect Zone and Min Dialect Zone



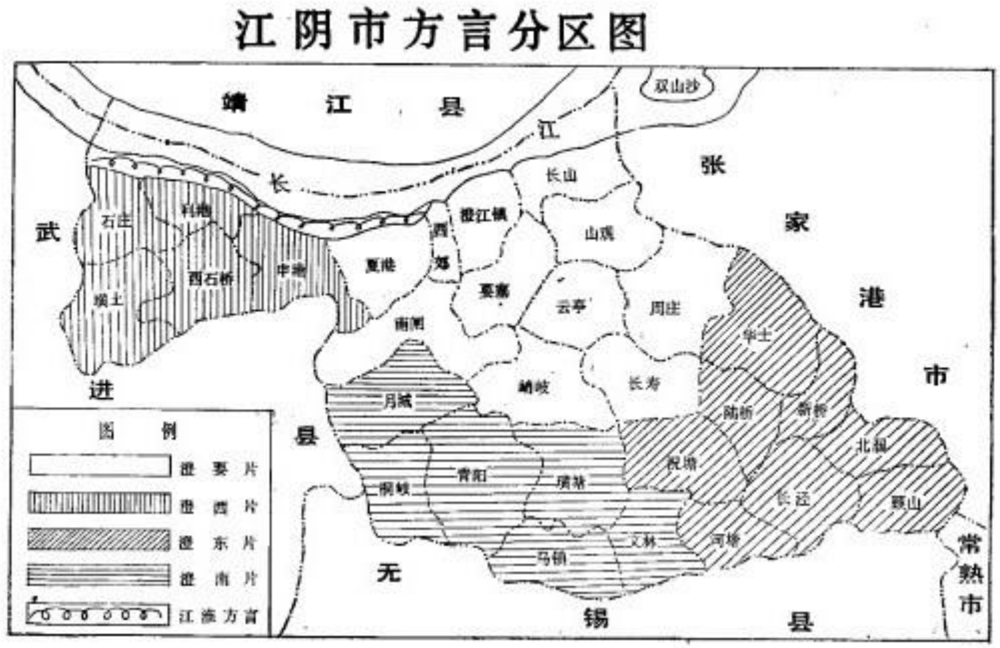
Notes: This figure shows the geographical distribution of the two dialect zones investigated by this study: Cantonese (Yue) dialect zone and Min dialect zone.

Figure 2: Share of Investment from Hong Kong and Taiwan



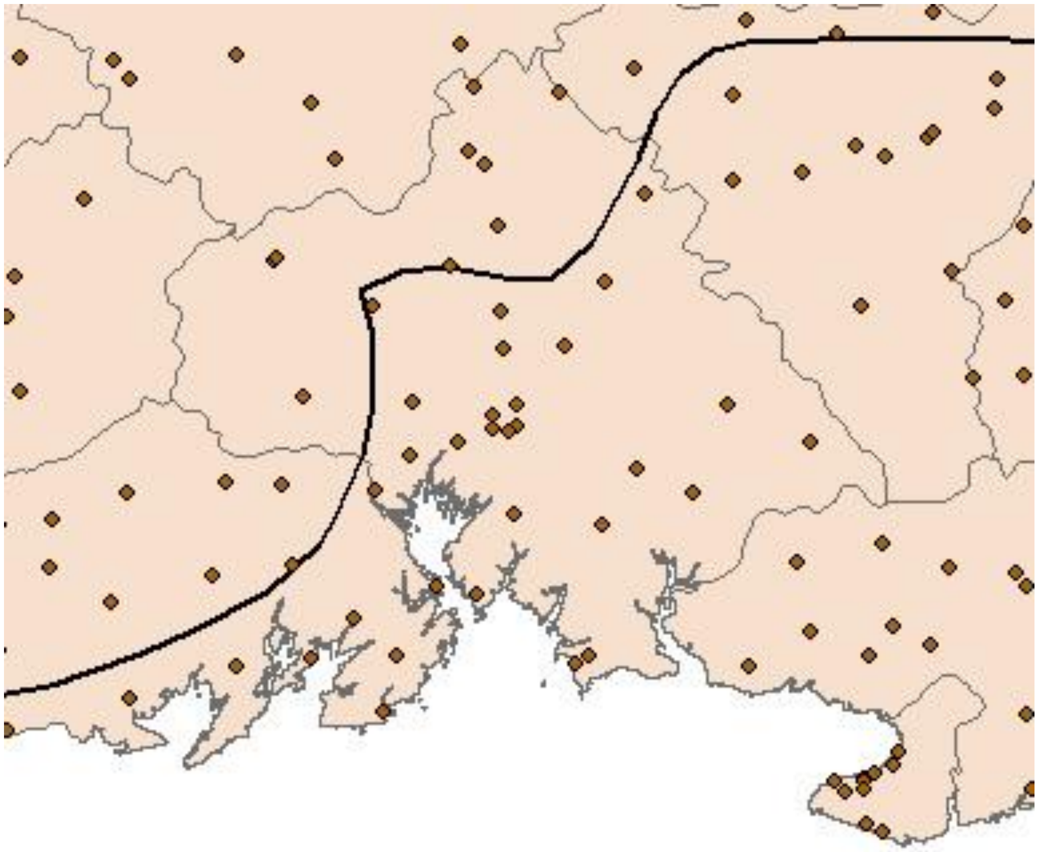
Notes: This figure shows the share of Hong Kong and Taiwan investment to mainland China among all foreign investments. Dashed line indicates share of Hong Kong investment while solid line indicates share of Tai wan investment. Data source: China statistical year books.

Figure 3: Example of Dialect Border



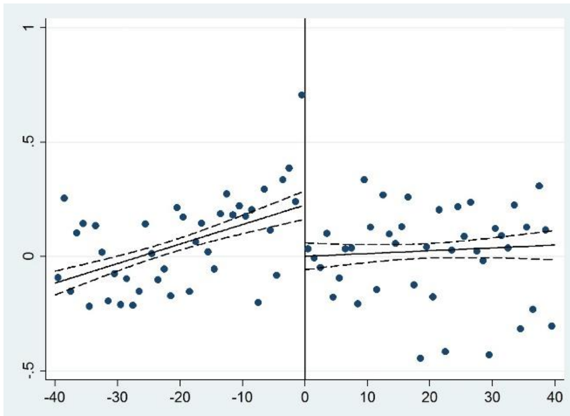
Notes: This figure shows an example of how dialect border is constructed from knowledge on major dialect used by villages of China. (The example county is not in Cantonese or Min dialect zone)

Figure 4: Dialect Border and Administrative County Border

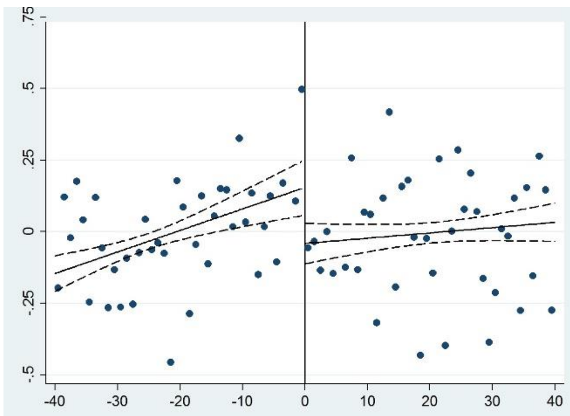


Notes: This figure shows an example of the relationship between dialect border and county border. The solid black line denotes Cantonese dialect border. The shallow grey line denotes administrative county border. The points refer to postcodes. We can see that an administrative county could be separated into two different dialect zones.

Figure 5: Discontinuity in Share of HMT firms at the border of dialect zone



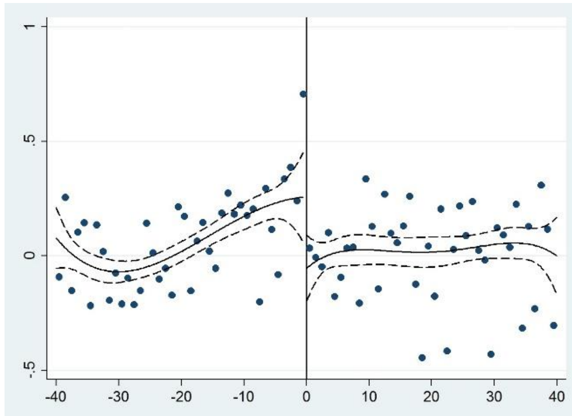
5.1. Employment Share



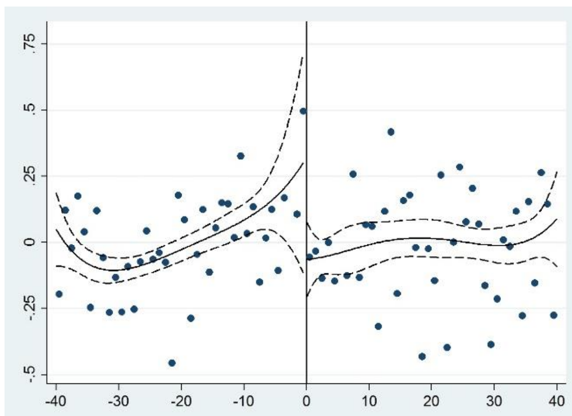
5.2. Output Share

Notes: These figures show the share of HMT firms among all firms at each postcode by distance to dialect border. All postcodes within 40 kilometers are included in the analysis. Horizontal axis denotes distance to dialect borders with negative value indicating the postcode is located inside the border (same dialect as HMT). Vertical axis denotes average share of HMT firms for a given distance. Figure 5.1 shows employment share; Figure 5.2 shows output share. Share of HMT firms is standardized by subtracting county mean and dividing by county standard deviation.

Figure 6: Discontinuity in Share of HMT firms at the border of dialect zones (Non-linear)



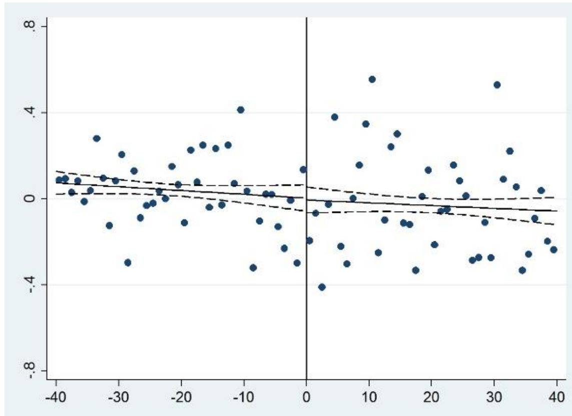
6.1. Employment Share



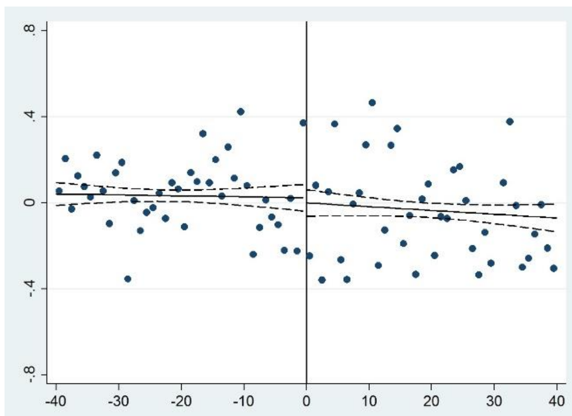
6.2 Output Share

Notes: These figures show the share of HMT firms among all firms at each postcode by distance to dialect border. All postcodes within 40 kilometers (40,000 meters) are included in the analysis. Horizontal axis denotes distance to dialect borders with negative value indicating the postcode is located inside the border (same dialect as HMT). Vertical axis denotes average share of HMT firms for a given distance. The model is fitted using 4th order of polynomials of distance to border. Figure 6.1 shows employment share; Figure 6.2 shows output share. Share of HMT firms is standardized by subtracting county mean and dividing by county standard deviation.

Figure 7: Placebo Tests on Share of firms from other countries at the border of dialect zones



7.1 Employment Share



7.2 Output Share

Notes: These figures show the share of other foreign firms among all firms at each postcode by distance to dialect border. These figures serve as placebo tests. Horizontal axis denotes distance to dialect borders with negative value indicating the postcode is located inside the border (same dialect as HMT). Vertical axis denotes average share of other foreign firms for a given distance. Figure 7.1 shows employment share; Figure 7.2 shows output share. Share of other foreign firms is standardized by subtracting county mean and dividing by county standard deviation.

Appendix 1: TFP estimation using non-parametric method.

Following Ackerberg et al. (2015), I use the following procedures to non-parametrically estimate labor and capital share (β_L and β_k) in the production function. Then I replace $\widetilde{S_{ft}}$ with β_L and $1-\widetilde{S_{ft}}$ with β_k in equation (6) to get TFP.

Suppose empirical production function is specified as $y = \beta_0 + \beta_l \times l + \beta_k \times k + \omega + \varepsilon$, where y denotes logarithm of value added, l denotes logarithm of labor input and k denotes logarithm of capital input. ω denotes TFP which is unobservable to the researcher. ε denotes random productivity shock.

I use the following procedures to get $\widehat{\beta_l}$ and $\widehat{\beta_k}$:

(1) Non-parametrically regress y on l, k and intermediary input and get predicted $\hat{y} = \hat{\varphi}(l, k, input)$;

(2) Then TFP ω can be estimated as $\widehat{\omega} = \hat{\varphi}(l, k, input) - \beta_0 - \beta_l \times l - \beta_k \times k$.

(3) Assuming for each firm, $\omega_t = \rho\omega_{t-1} + u_t$, where u_t denotes exogenous productivity shock, given parameters ρ, β_0, β_l and β_k, $\widehat{\omega}_t$, $\widehat{\omega}_{t-1}$ and $\widehat{u_t}$ can be calculated from the data.

(4) Using the moment condition that $\widehat{u_t}$ is orthogonal to $1, l_{t-1}$, k_t and $\hat{\varphi}(l, k, input)_{t-1}$ and GMM to estimate $\widehat{\beta_l}$ and $\widehat{\beta_k}$.

Using sample of firms within 40 kilometers of the dialect borders, I estimate that $\widehat{\beta_l} = 0.2793$ and $\widehat{\beta_k} = 0.4429$.