Place-based Land Policy and Firm Productivity:
Evidence from China's Eastern-Inland Regional Border*

Min Fang — UF
Libin Han — DUFE
Zibin Huang — SUFE
Ming Lu — SJTU
Li Zhang — SYSU

March 27, 2024
latest version

Abstract

We study the effect of China's inland-favoring land policy on firm-level productivity by employing a research design combining difference-in-differences and regression discontinuity at the policy border. We find that the inland-favoring land policy decreased the firm productivity gap between developed (eastern) regions and underdeveloped (inland) regions. The relative changes are mainly due to slower eastern firm productivity growth rather than faster inland firm productivity growth. Eastern firms reduced their R&D expenditure and capital usage as a response to the policy.

Keywords: Place-based Policy; Land Policy; Spatial Misallocation; Migration; Labor Mobility; Regional Inequality; China; **JEL Classification Numbers:** O18, R58, E24, J61, R52;

*Min Fang (minfang@ufl.edu), Department of Economics, University of Florida; Libin Han (hanlibin@126.com), Economic and Social Development Institute, Dongbei University of Finance and Economics; Zibin Huang (huangzibin@mail.shufe.edu.cn), College of Business, Shanghai University of Finance and Economics, and Shanghai Institute of International Finance and Economics; Ming Lu (luming1973@sjtu.edu.cn), Antai College of Economics and Management, Shanghai Jiao Tong University; Li Zhang (zhangl39@mail.sysu.edu.cn), International School of Business and Finance, Sun Yat-Sen University.

This note was previously part of the seventh version of Fang et al. (2022) (https://doi.org/10.5281/zenodo.10126607) but is now revised to serve as a companion piece for Fang et al. (2022). We thank audiences at numerous seminars and conferences for their helpful comments. All errors are ours.

1 Introduction

In this paper, we empirically study the consequences of a major place-based land allocation policy change on firm-level productivity in China. Specifically, we investigate a sudden shift in China's land supply policy in 2003, commonly called the inland-favoring land policy. The allocation of construction land quotas has been used as a place-based policy since 2003. Before 2003, developed areas with higher land demand were usually assigned more land quota. However, since 2003, the central government focused on balancing economic development by allocating more land quota to underdeveloped inland provinces. In this paper, we aim to find the direct causal effect of this policy change on the measured productivity of firms.

In Fang et al. (2022), we investigate the impact of the inland-favoring land supply policy on TFP at the prefecture level. By aggregating firm-level TFP across different prefectures, weighted according to firm employment, we conduct a traditional prefecture-level Difference-in-Differences regression analysis. In this supplementary note, we further substantiate our empirical findings by directly investigating the policy's effects at the firm level.

A typical identification problem is that firms in the eastern region are usually very different from those in other regions in terms of both observed and unobserved characteristics. To solve this endogeneity issue, we employ a method combining border Regression Discontinuity Design (Black, 1999) and the Difference-in-Differences approach (RD-DID). The basic idea is that firms within a minimal bandwidth along the border are very similar, no matter which side they are located on. Thus, firm-level TFP should have similar time trends. This allows us to implement an RD-DID strategy on these border firms to identify the effects of the inland-favoring land policy at the firm level. Compared with the prefecture-level regression, the advantage of this firm-level regression is that it exploits more variation and provides detailed micro evidence.

We show that the inland-favoring policy reduced the firm-level TFP gap between the eastern and inland regions by approximately 8%, which is consistent with the results at the prefecture level. The results remain robust across various robustness exercises in our regression analysis. Moreover, we do not observe significant TFP improvements among inland firms. Our empirical analysis demonstrates that the inland-favoring land policy narrows the productivity gap between eastern and inland firms by adversely affecting eastern firms without significantly benefiting inland firms, suggesting that land constraints could be a potential cause.

There are two potential channels for firm-level TFP gap changes. First, a more restrictive land policy in eastern regions could precipitate the exit of lower productivity firms from the market or compel them to downscale their operations below the survey threshold of the NIED. This phenomenon is referred to as the "selection effect." Second, the policy may have directly impeded the productivity of eastern firms by constraining their land usage, slowing their technology innovation, and diminishing their agglomeration benefits. We term this the "direct effect". We investigate these two channels separately and find that the direct effect plays an important role. The inland-favoring land policy significantly hindered firm innovation and reduced R&D expenditures. The capital usage is also decreased. On the contrary, we find no evidence for the selection effect.

2 Data

We use the National Industrial Enterprise Database (NIED), published by the National Bureau of Statistics. It covers all state- and non-state-owned enterprises "above scale" (main business revenue greater than 5 million RMB). This dataset accounts for over 90% of all industrial production in China.[1] The dataset contains rich enterprise-level information, such as firm name, four-digit industry category, incorporation year, number of employees, total salary, and total fixed assets.[2] Table 1 shows the descriptive statistics of the enterprise data. Our main TFP calculation is based on the OP (Olley and Pakes, 1992) estimation method. We also calculate TFP using the LP (Levinsohn and Petrin, 2003) method in Appendix A, which yields similar results.

Table 1: **Summary Statistics**

Variable	Description	Observations	Mean	Std. dev.	Min	Median	Max
Ln(tfp_op)	TFP(OP)	877383	3.25	1.02	-0.04	3.27	5.63
Ln(tfp_lp)	TFP(LP)	877383	6.36	1.09	3.08	6.32	9.02
Ln(output)	Ln(1k yuan)	877383	8.62	1.29	5.31	8.51	12.22
Ln(wage)	Ln(1k yuan)	876147	2.39	0.63	0.39	2.41	4.14
Age	Year	877383	9.66	9.22	1	7	48
Employee	Person	877383	192.37	293.80	12	97	1985
East	Dummy	877383	0.80	0.40	0	1	1
Firm Distance	Km	877383	76.06	102.32	-199.99	102.52	200

Notes: East is a dummy variable set to 1 if the firm is in the eastern area. Firm distance is from the firm's location to the east–inland provincial boundary, which is positive for eastern firms and negative for inland firms. All chosen observations are within 200 km of the boundary.

3 Main Empirical Analysis

We empirically analyze how the inland-favoring land supply policy affected firm performance, emphasizing the effects on firm-level TFP. We show causal evidence that this policy shrank the

[1] Since there is a major missing data issue after 2007, we only use data from 1998 to 2007.

[2] For unknown reasons, some companies provide missing or erroneous information. Some data cleaning and a 1% censoring process were applied to avoid abnormal observations.

TFP gap between eastern and inland firms. This reduction in the gap can be primarily attributed to the decreased TFP of eastern firms.

3.1 Empirical Specification

The main empirical strategy in analyzing firm TFP combines a Border Regression Discontinuity Design as in Black (1999) and a Difference-in-Differences approach (RD-DID). The basic idea is to first compare firm TFP on the eastern and inland sides of the border. Then, we compare this border TFP difference over time, particularly before and after the year when the central government implemented the inland-favoring land supply policy. If the time trend of TFP is similar in the neighborhood of the border, the DID design can identify the policy effect. Figure 1 shows the location of the boundary between the eastern and inland regions of China. Red dots represent firms on the eastern side of the boundary. Black dots represent firms on the inland side of the boundary. We use the region definitions published by the National Bureau of Statistics of China.[3]

For firm i at border segment b in city c and year t, we have the following regression:

$$
\begin{aligned}
ln(y_{ibct}) = \alpha &+ \beta_1 East_{ibt} + \beta_2 f(Dist_{ibt}) + \beta_3 East_{ibt} \times f(Dist_{ibt}) \\
&+ Post2003 \times [\delta_1 East_{ibt} + \delta_2 f(Dist_{ibt}) + \delta_3 East_{ibt} \times f(Dist_{ibt})] \\
&+ \beta_4 X_{ct-1} + \phi_{bt} + \gamma_t + \psi_i + \epsilon_{ibct} \qquad (1)
\end{aligned}
$$

where y_{ibct} is the log TFP of firm i. $East_{ibt}$ is a dummy that equals one if the firm is located on the eastern side of the border, which carries a subscript t since firms can change their locations across time. $f(Dist_{ibt})$ is a smooth function of the distance between the firm and the border, and $Post2003$ is a dummy which equals one if t is after 2003 (including 2003 itself).[4] X_{ct-1} is a set of lagged city-level control variables, including the log of GDP, the log of population, the log of city area, and the value added to the service sector. ϕ_{bt} is the border segment fixed effect for the firm at time t. We divide the border into five segments of equal length and designate each firm to the nearest segment. γ_t is the year fixed effect. ψ_i is the firm fixed effect.[5]

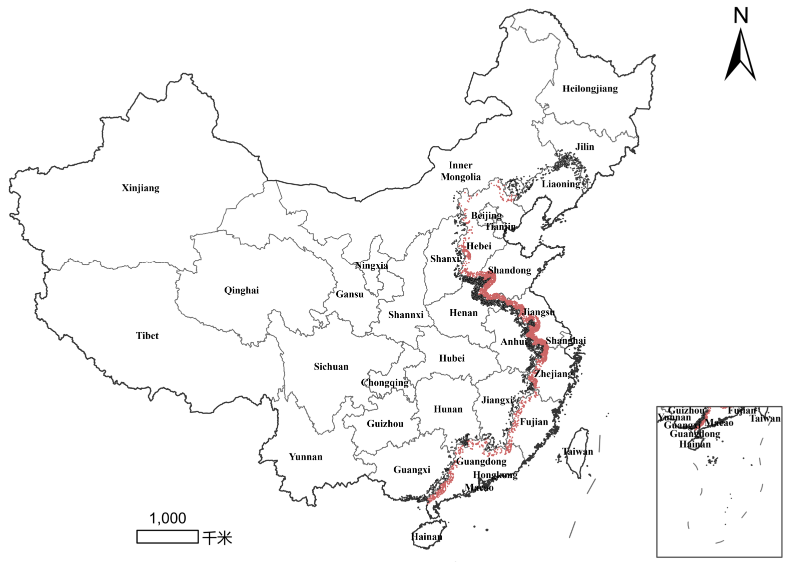
This is a regression combining RD and DID methods. First, consider the first three terms (except the intercept), that is, $\beta_1 East_{ibt} + \beta_2 f(Dist_{ibt}) + \beta_3 East_{ibt} \times f(Dist_{ibt})$. This comprises a border regression discontinuity design regression, with the running variable being the distance to the border. Using only the observations within a small bandwidth, we assume that firms just on the eastern side of the border are very similar to firms just on the inland side. By fitting a smooth function $f(Dist)$, β_1 captures the effect of being in the eastern region on outcome variable y. This

[3]We consider northeastern provinces as inland.

[4]We also run all regressions in a specification where 2003 is excluded from the treatment group. The results are not qualitatively changed.

[5]We also investigate a simpler regression setting without a firm fixed effect. The results hold qualitatively.

Figure 1: **China's Eastern-Inland Boundary**



Notes: The boundary is between eastern provinces and their inland neighbors. Red dots represent firms on the eastern side of the boundary. Black dots represent firms on the inland side of the boundary (To avoid confusion, the black dots on the eastern coastline are just islands, which are not part of our firm sample.). The data source is the National Bureau of Statistics of China.

study uses two fitting functions: local linear regression and linear regression.

Second, we add interactions between the post-2003 dummy and all previous RD terms. Coefficient δ_1 then denotes the policy effect, which is interpreted as the change in the eastern region's TFP premium over the inland region before and after the 2003 inland-favoring land allocation policy. This is a difference-in-differences estimation. The first difference is between the eastern and inland regions (at the border, within the bandwidth). The second difference is between the before-policy (2003) and the after-policy periods. In general, this specification combines border regression discontinuity design with difference-in-differences.

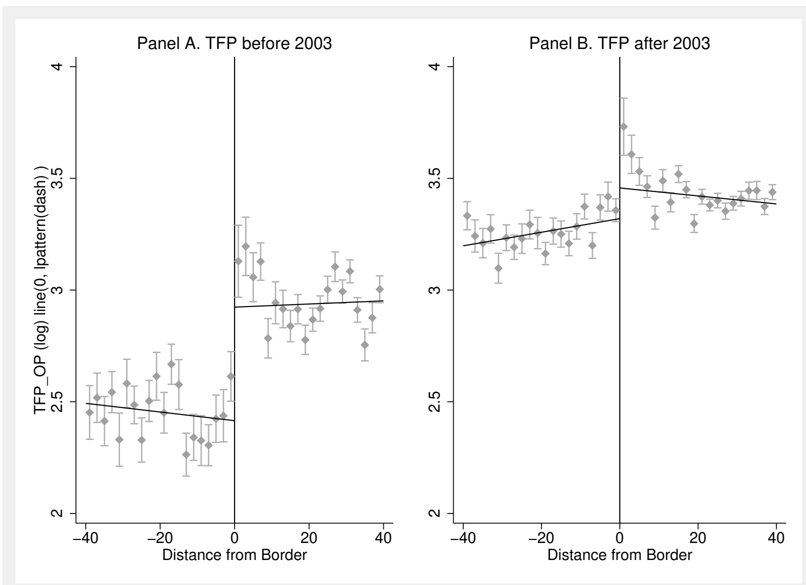
It is important to clarify that the inland-favoring land policy can potentially affect the TFP levels of both regions. Therefore, the regression coefficient should be interpreted as the policy's effect on the regional gap (relative level) rather than on the absolute level of TFP for either region.

3.2 Regression Assumptions Validation

We validate our regression method by checking several important assumptions.

First, we investigate the existence of the boundary discontinuity by drawing an RD figure. Figure 2 depicts panel A, representing data before 2003, and panel B, representing data after 2003. The x-axis displays the distance of firms from the boundary, with a positive distance indicating firms located on the eastern side. The y-axis displays firm-level TFP, calculated using the Olley and Pakes (1992) method. This reveals a distinct discontinuity along the eastern-inland border in both panels. Notably, this gap narrowed following the implementation of the 2003 inland-favoring land policy.

Figure 2: **Regression Discontinuity Changes**

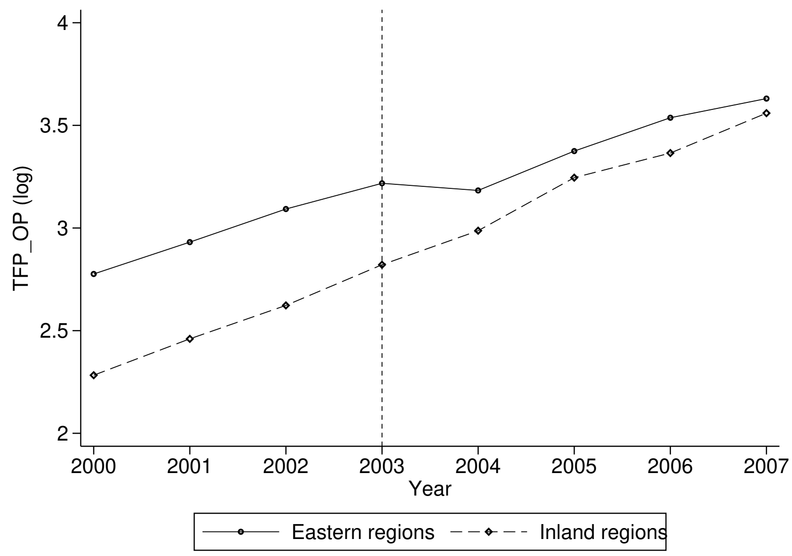


Notes: The dependent variable is firm-level TFP calculated using the Olley and Pakes (1992) method. The smoothing function is linear. The bandwidth is 40 km from the border.

Second, we investigate the eastern and inland time trends of firm TFP. Our regression specification assumes that firm parcels on the eastern and inland sides of the border should have a similar time trend. Figure 3 shows the time trends of firm-level TFP. The black line is average TFP in the developed eastern region, and the grey line is average inland TFP. The dashed vertical line is located just after 2003 when the inland-favoring land policy was implemented. We find no evidence of divergent time trends in firm productivity before the policy. Despite the 2003 policy's aim to boost inland development, we do not observe a corresponding increase in the growth rate of inland TFP. Instead, the policy seems to have stymied the growth of eastern TFP.

Finally, we implement a traditional event study regression to investigate the evolution of the eastern region effect across time. We take 2003 as the baseline year and then run the following

Figure 3: **Time Trends of Firm TFP**



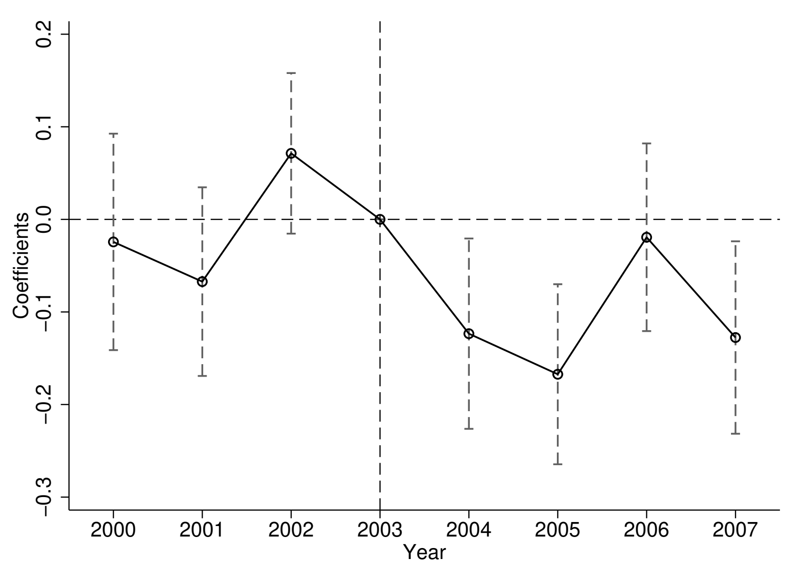
Notes: This figure shows the time trends of firm-level TFP calculated using the Olley and Pakes (1992) method and land parcel prices. The black line is average TFP in the developed eastern region, and the grey line is average inland TFP. The dashed vertical line indicates the implementation of the inland-favoring land policy. TFP is calculated using only firms within 40km of the border. Firm-level TFP and land prices followed similar trends before the policy.

regression for the event study:

$$\begin{aligned} ln(y_{ibct}) &= \alpha + \beta_1 East_{ibt} + \beta_2 f(Dist_{ibt}) + \beta_3 East_{ibt} \times f(Dist_{ibt}) \\ &+ \sum_{s \neq 2003} \mathbf{1}(s = t) \times [\delta_{1s} East_{ibt} + \delta_{2s} f(Dist_{ibt}) + \delta_{3s} East_{ibt} \times f(Dist_{ibt})] \\ &+ \beta_4 X_{ct-1} + \phi_b + \gamma_t + \psi_i + \epsilon_{ibct} \end{aligned} \quad (2)$$

We plot the evolution of the coefficient δ_{1s} across time s in Figure 4, illustrating the changing of the eastern region effect across time, with 95% confidence intervals. We choose a linear smoothing function. We find that all the coefficients are very close to zero before 2003. They became statistically and economically distinguishable from zero only after implementing the policy. The results from this event study confirm that there is no pre-trend in our data. These figures also give us a preview of the main results. After the central government imposed the inland-favoring land policy in 2003, there was a relative decrease in the firm productivity gap between the eastern and inland regions.

Figure 4: **Event Study - TFP (OP)**



Notes: The dependent variable is firm-level TFP calculated using the Olley and Pakes (1992) method. The bandwidth is 40 km from the border. The corresponding confidence interval is 95%.

3.3 Empirical Results

Main Results Table 2 shows the regression results based on TFP. In the two columns, we use a local and linear fit for the smoothing function, respectively. We use the optimal bandwidth for the local linear fit based on Imbens and Kalyanaraman (2012). The bandwidth we use for the linear fit is 40 km.[6] In the first column, we use local linear regression as our fitting function. In the second column, we change the fitting function to be a global first-order polynomial (linear). We find that the reduction in land supply after 2003 reduced the measured TFP gap of eastern firms relative to inland firms by about 8%.

Robustness Checks We also implement nine groups of robustness analyses to address an extensive set of potential empirical concerns. The results are available in Appendix A.

The first group addresses concerns with the robustness of our TFP estimates. We verify robustness by conducting the empirical analysis using firm-level TFP calculated with the methods proposed by Levinsohn and Petrin (2003). Table A1 shows that the results are very similar to the main results. The second group addresses concerns with the robustness of our bandwidth choice. We vary the bandwidths for the linear and quadratic smoothing functions between 20 and 70 km in Tables A2 and A3. The results are very robust qualitatively. The third group addresses concerns with potential bad control issues. We run all main regressions without city-level

[6] We also try other bandwidths, and the results are similar. Please refer to Appendix A for details.

Table 2: **RD-DID Results on TFP (OP)**

	(1) Local Linear	(2) Poly RD (Poly=1)
Post2003×East	-0.0798**	-0.0761*
	(0.0356)	(0.0426)
City Lagged Controls	Y	Y
Border FE	Y	Y
Year FE	Y	Y
Firm FE	Y	Y
Observations	131,250	100,054
R-squared	0.1203	0.1161

Notes: The dependent variable is firm-level TFP measured by the Olley and Pakes (1992) method. The set of lagged city-level control variables includes the log of GDP, the log of population, the log of city area, and the scale of the service sector. The sample in the local linear regression specification is restricted to be within an optimal bandwidth using a constant kernel. The sample in the polynomial RD case is restricted to be within a bandwidth of 40 km around the raw boundary. The standard errors are clustered at the firm level. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

lagged control variables to address any potential bad control issues. Tables A4 and A5 show that the resulting estimates are similar to those with control variables. Both the point estimates and R-squares exhibit minimal changes, validating our regression results according to Oster (2019).

In the fourth group, we simplify the regression discontinuity functional form by keeping slopes unchanged at the boundary. Table A6 shows minimal change compared with the baseline results. In the fifth group, we alleviate potential contamination from special geographical characteristics at the provincial boundary by excluding firms within 10 km on either side. Table A7 shows that the results have not changed. In the sixth group, we investigate the effect of firms moving their locations. In Figure A1 and Table A8, we show that the number of relocating firms is minimal, and no regression results change if we drop these firms. This is reasonable since the National Industrial Enterprise Database firms are all "above scale" large firms that rarely change their locations. In Table A9, we perform a placebo test by shifting the boundary to the west or the east. We do not observe significant changes in the gaps for these artificial boundaries before or after 2003.

We also address concerns about possible confounders around 2003. In the seventh group of robustness checks, we address the potential spatial effect of China joining the WTO in 2001. To address this issue, we run regressions keeping only firms with zero exports and regressions controlling for firm-level exports to eliminate any WTO effect. The regression results in Tables A11, A12, A13 and A14 show that the main conclusions are unchanged. In the eighth group, we try to rule out the effects of some other subsidy and tax policies happening concurrently, which may distort our estimates. Tables A15, A16, and A17 show that the main results are maintained.

3.4 Remarks on Main Results

We show that the inland-favoring land policy decreased the firm productivity gap between developed eastern regions and underdeveloped inland regions. Based on the time trends of firm TFP in Figure 3, the relative changes are mainly due to the reduction in eastern firm productivity growth rather than improving inland productivity. These findings indicate that although government achieved the goal of shrinking the eastern-inland gap, it potentially came at a substantial cost of distorting land prices and decreasing the productivity of eastern firms. In other words, such regional convergence comes at the cost of spatial misallocation. In the next section, we will investigate the mechanism at the firm level in more detail.

4 Selection Effect or Direct Effect

At the firm level, the policy influences TFP through two distinct channels. First, a more restrictive land policy in eastern regions could precipitate the exit of lower productivity firms from the market or compel them to downscale their operations below the survey threshold of the NIED. This could, in turn, elevate the average TFP of the location, a process we refer to as the "selection effect." Second, such a policy could directly damage the productivity of existing eastern firms, for example through a reduction in their land inputs, increased production costs, or decreased R&D expenditure, and a consequent decline in regional agglomeration. We label this the "direct effect." In this section, we separately explore these two channels.

4.1 Selection Effect

In the NIED dataset, we classify firms that are present in year t but absent from the survey in year $t + 1$ as exiting firms. Conversely, firms that are present in both years are categorized as surviving firms. Figure 5 illustrates the proportion of exiting firms relative to the total number of firms for each year. On average, the exit rate fluctuates between 10% and 20%, with inland firms more likely to exit. Notably, there was a significant spike in 2003, where the exit rate for firms in both regions escalated to over 25%.

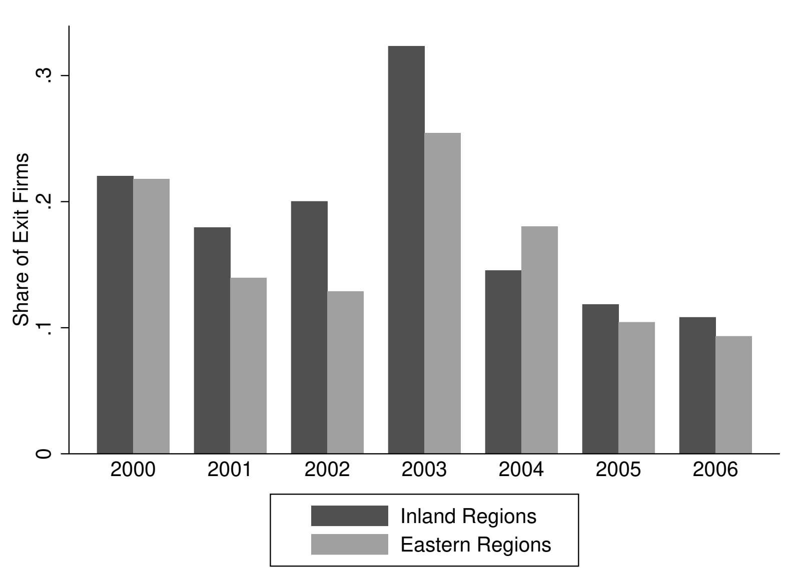
Figure 6 shows the average TFP for exiting firms (red solid line) and surviving firms (blue dashed line) across years. Subfigure (a) illustrates inland firms, and subfigure (b) illustrates firms in eastern coastal regions. We find that the TFP gap between exiting and surviving firms shrunk after 2003 in both inland and eastern regions. The gap was reduced more in the east than inland. There is no evidence in our data supporting a positive selection effect among firms after the 2003 inland-favoring land supply policy. The productivity gap between exiting and surviving firms is smaller in eastern regions than inland. Thus, the reduction of eastern relative TFP cannot be explained by the exit of low productivity firms. In Figures 7 and 8, we further investigate the trends for averages of firms' total assets and employment. We detect no evidence for a larger positive selection of surviving firms in eastern regions after 2003.

To precisely estimate the changes in selective pressure for firm i in year t, we run the following Difference-in-Differences-in-Differences (DDD) regression:

$$\begin{aligned} y_{it} = \beta_0 + & \beta_1 Exit_{it} \times Post2003_t \times East_j + \beta_2 Exit_{it} \times Post2003_t + \beta_3 Post2003_t \times East_j \\ & + \beta_4 Exit_{it} \times East_j + \phi_i + \gamma_t + \beta_5 Exit_{it} + \epsilon_{it} \end{aligned} \quad (3)$$

$Exit_{it}$ indicates whether firm i in year t is an exiting firm that will disappear in the next year's survey. β_1 is the parameter of interest, which evaluates whether the eastern-inland productivity/asset/employment gap between exiting and surviving firms changed after 2003. It represents the effect of the inland-favoring policy on firm selection. If we detect a significantly negative coefficient, it means that this policy led to more selection and drove firms with low productivity out of market in eastern regions, causing the remaining firms to have higher productivity than the exiting firms. Table 3 shows the results of this DDD regression, and we do not find any evidence for discrepancies in selection across regions.

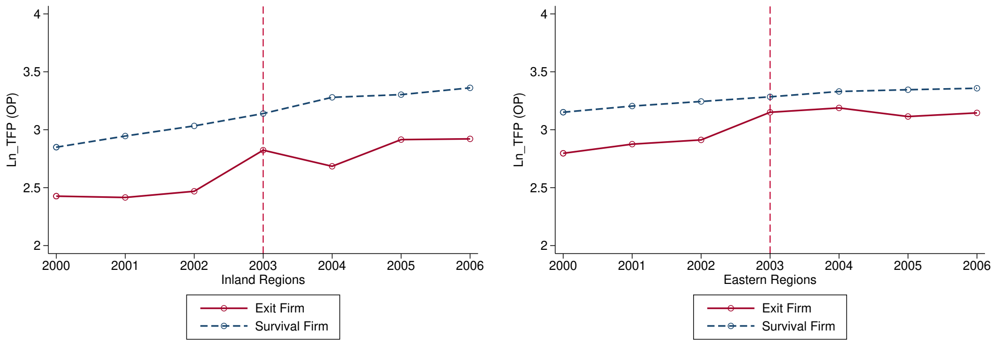
Figure 5: **Exit Rate of Firms from NIED**



Notes: This figure shows the firm exit rate from the NIED Survey each year.

Figure 6: TFP (OP) of Exiting and Surviving Firms by Regions

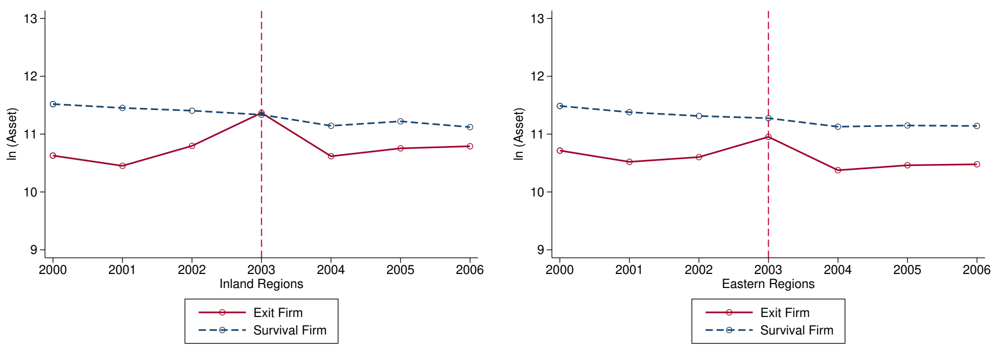
(a) Inland Firms' TFP (b) Eastern Firms' TFP



Notes: The data source is the National Industrial Enterprise Database. The blue dashed line represents surviving firms. The red solid line represents exiting firms. Subfigure (a) shows the TFP changes for inland firms. Subfigure (b) shows the TFP changes for eastern coastal firms.

Figure 7: Total Assets of Exiting and Surviving Firms by Region

(a) Inland Firms' Total Assets (b) Eastern Firms' Total Assets



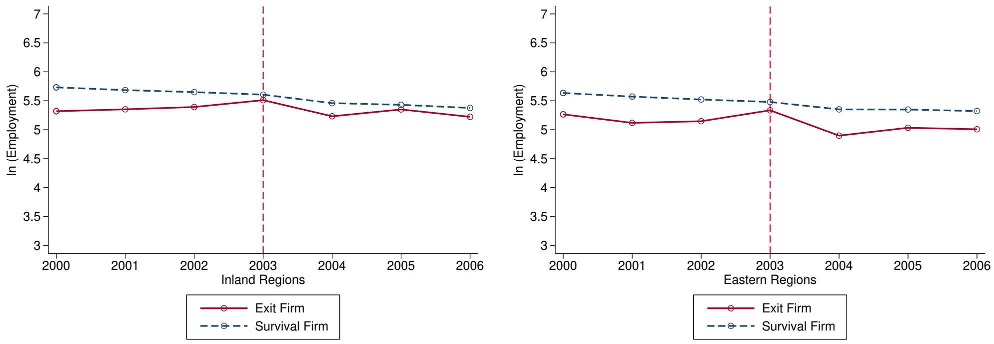
Notes: The data source is the National Industrial Enterprise Database. The blue dashed line represents surviving firms. The red solid line represents exiting firms. Subfigure (a) shows the changes in total assets for inland firms. Subfigure (b) shows the changes in total assets for eastern coastal firms.

4.2 Direct Effect

In the last section, we have shown that selection is not a significant part of our story. Now we investigate the direct effect by running the same main regression but using firms' factor inputs and total outputs as the dependent variables. In Table 4, we consider five variables: return on assets in columns (1) and (2); log of labor input (employment) in columns (3) and (4); log of capital input in columns (5) and (6); log of total output in columns (7) and (8); log of R&D expenditure in columns (9) and (10). In the odd columns, we use a local linear smoothing function. In the even

Figure 8: Employment of Exiting and Surviving Firms by Regions

(a) Inland Firms' Employment (b) Eastern Firms' Employment



Notes: The data source is the National Industrial Enterprise Database. The blue dashed line represents surviving firms. The red solid line represents exiting firms. Subfigure (a) shows the average employment changes for inland firms. Subfigure (b) shows the average employment changes for eastern coastal firms.

Table 3: **DDD Results**

	(1) TFP	(2) Total Assets	(3) Employment
Exit×Post2003×East	-0.0152	0.0189*	-0.0072
	(0.0183)	(0.0105)	(0.0110)
Double Interactions	Y	Y	Y
Exiting Dummy	Y	Y	Y
Year FE	Y	Y	Y
Firm FE	Y	Y	Y
Observations	805,906	805,906	805,906
R-squared	0.7061	0.9364	0.9047

Notes: The dependent variables are firm-level TFP measured by the Olley and Pakes (1992) method, firms' total assets, and firms' employment. The standard errors are clustered at the firm level. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

columns, we use a linear smoothing function.

We find that after 2003, firms in eastern regions experienced reductions in relative output and R&D expenditures compared to their inland counterparts. Additionally, the relative capital input of eastern firms in declined, albeit to a lesser extent and without statistical significance. No significant effects were observed for return on assets and labor input. Consequently, we deduce that the primary factor behind the decrease in the relative TFP of eastern firms is the reduction in their R&D expenditures. This decline could naturally stem from diminished agglomeration, where the knowledge spillovers were restricted by the land-use policy.

Table 4: **RD-DID Results on Firm Inputs**

	ROA		Ln(Labor)		Ln(Capital)		Ln(Output)		Ln(R&D)	
	(1) LL	(2) Poly	(3) LL	(4) Poly	(5) LL	(6) Poly	(7) LL	(8) Poly	(9) LL	(10) Poly
Post2003×East	0.0132	0.0057	0.0229	0.0327	-0.0384	-0.0655	-0.0495	-0.0855*	-0.1585**	-0.0906*
	(0.0110)	(0.0064)	(0.0715)	(0.0280)	(0.0649)	(0.0476)	(0.0778)	(0.0468)	(0.0763)	(0.0532)
City Lagged Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Border FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Firm FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	36,606	100,054	15,363	100,054	57,514	100,054	34,371	100,054	55,234	100,053
R-squared	0.0603	0.0475	0.0155	0.0125	0.0583	0.0547	0.1977	0.1736	0.0745	0.0779

Notes: The dependent variables are return on assets, labor input, capital input, total output, and R&D expenditure. The set of lagged city-level control variables includes the log of GDP, the log of population, the log of city area, and the scale of the service sector. The sample in the local linear regression specification is restricted to be within an optimal bandwidth using a constant kernel. The sample in the polynomial RD cases is restricted to be within a bandwidth of 40 km around the raw boundary. The standard errors are clustered at the firm level. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

5 Conclusion

In this note, we provide direct causal evidence of how China's 2003 inland-favoring land policy affected firm-level productivity by examining firms at the border of the eastern and inland regions. The inland-favoring land policy decreased the firm productivity gap between the more developed eastern regions and the underdeveloped inland regions. Furthermore, the relative changes are mainly due to the reduction in eastern firm productivity growth rather than growth in inland firm productivity. These findings indicate that although government may have achieved the goal of shrinking the regional gap, it potentially came at the cost of distorting land and housing prices and reducing eastern firms' productivity.

In Fang et al. (2022), we leverage these findings to city-level empirical analysis and a prefecture-level spatial general equilibrium to show the aggregate effects. The quantitative analysis shows that, indeed, although the government achieved the goal of shrinking the regional gap between the eastern and inland regions, it came at a substantial cost of more severe spatial misallocation of production and labor. Meanwhile, as shown in both Fang et al. (2022) and Fang and Huang (2022), such place-based land policy may actually increase income inequality among workers because those from underdeveloped regions are disincentivized to migrate to more developed regions offering high wages because of expensive housing.

References

Abadie, Alberto, Susan Athey, Guido W Imbens, and Jeffrey M Wooldridge. 2020. "Sampling-based versus Design-based Uncertainty in Regression Analysis." *Econometrica* 88 (1):265–296.

———. 2023. "When Should You Adjust Standard Errors for Clustering?" *The Quarterly Journal of Economics* 138 (1):1–35.

Angrist, Joshua D and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton University Press.

Bernard, Andrew B, J Bradford Jensen, Stephen J Redding, and Peter K Schott. 2007. "Firms in International Trade." *Journal of Economic Perspectives* 21 (3):105–130.

———. 2018. "Global Firms." *Journal of Economic Literature* 56 (2):565–619.

Black, Sandra E. 1999. "Do Better Schools Matter? Parental Valuation of Elementary Education." *The Quarterly Journal of Economics* 114 (2):577–599.

Fang, Min, Libin Han, Zibin Huang, Ming Lu, and Li Zhang. 2022. "Place-based Land Policy and Spatial Misallocation: Theory and Evidence from China." *Available at SSRN 3846313* .

Fang, Min and Zibin Huang. 2022. "Migration, Housing Constraints, and Inequality: A Quantitative Analysis of China." *Labour Economics* 78:102200.

Imbens, Guido and Karthik Kalyanaraman. 2012. "Optimal Bandwidth Choice for the Regression Discontinuity Estimator." *The Review of Economic Studies* 79 (3):933–959.

Levinsohn, James and Amil Petrin. 2003. "Estimating Production Functions Using Inputs to Control for Unobservables." *The Review of Economic Studies* 70 (2):317–341.

Michalopoulos, Stelios and Elias Papaioannou. 2014. "National Institutions and Subnational Development in Africa." *The Quarterly Journal of Economics* 129 (1):151–213.

Olley, G Steven and Ariel Pakes. 1992. "The Dynamics of Productivity in the Telecommunications Equipment Industry." Tech. rep., National Bureau of Economic Research.

Oster, Emily. 2019. "Unobservable Selection and Coefficient Stability: Theory and Evidence." *Journal of Business & Economic Statistics* 37 (2):187–204.

A Appendix

In this section, we implement nine groups of robustness checks for our empirical analysis. We also investigate the policy effect on other outcome variables in the last subsection.

A.1 Robustness Checks for TFP Estimation Method

First, we implement the empirical analysis using firm-level TFP calculated using the method proposed by Levinsohn and Petrin (2003). Table A1 shows the results of the primary regression using this productivity measure instead of the OP method, and all results remain very similar.

Table A1: Robustness: RD-DID Results on TFP (LP)

	(1) Local Linear	(2) Poly RD (Poly=1)
Post2003×East	-0.0533	-0.0921**
	(0.0478)	(0.0439)
City Lagged Controls	Y	Y
Border FE	Y	Y
Year FE	Y	Y
Firm FE	Y	Y
Observations	85,748	100,054
R-squared	0.1416	0.1493

Notes: The dependent variable is firm-level TFP measured by the Levinsohn and Petrin (2003) method. The regression specifications are identical to Table 2. The standard errors are clustered at the firm level. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

A.2 Robustness Checks for Bandwidth Choices

Second, we adjust the bandwidth for the linear and quadratic smoothing functions. We present results for bandwidth choices ranging from 20 km to 70 km in Tables A2 and A3. The results remain qualitatively robust, although we lose observations when we reduce the bandwidth, leading to decreased estimation precision.

Table A2: Robustness: TFP Regressions with Different Bandwidth Choices (OP)

bandwidth	(1) 20km	(2) 30km	(3) 40km	(4) 50km	(5) 60km	(6) 70km
Post2003×east	-0.0235	-0.0076	-0.0761*	-0.0824**	-0.0574*	-0.0271
	(0.0682)	(0.0512)	(0.0426)	(0.0363)	(0.0330)	(0.0298)
City Lagged Controls	Y	Y	Y	Y	Y	Y
Border FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
Firm FE	Y	Y	Y	Y	Y	Y
Observations	39,747	72,488	100,054	126,265	152,064	184,678
R-squared	0.1303	0.1113	0.1161	0.1195	0.1208	0.1161

Notes: The dependent variable is firm-level TFP measured by the Olley and Pakes (1992) method. The set of lagged city-level control variables includes the log of GDP, the log of population, the log of city area, and the scale of the service sector. We use a linear fit as the smoothing function. The standard errors are clustered at the firm level. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

Table A3: Robustness: TFP Regressions with Different Bandwidth Choices (LP)

bandwidth	(1) 20km	(2) 30km	(3) 40km	(4) 50km	(5) 60km	(6) 70km
Post2003×east	-0.0056	-0.0041	-0.0921**	-0.0946**	-0.0688**	-0.0375
	(0.0694)	(0.0527)	(0.0439)	(0.0373)	(0.0341)	(0.0309)
City Lagged Controls	Y	Y	Y	Y	Y	Y
Border FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
Firm FE	Y	Y	Y	Y	Y	Y
Observations	39,747	72,488	100,054	126,265	152,064	184,678
R-squared	0.1644	0.1442	0.1493	0.1531	0.1547	0.1504

Notes: The dependent variable is firm-level TFP measured by the Levinsohn and Petrin (2003) method. The set of lagged city-level control variables includes the log of GDP, the log of population, the log of city area, and the scale of the service sector. We use a linear fit as the smoothing function. The standard errors are clustered at the firm level. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

A.3 Robustness Checks for Without City-level Controls

Third, we run all main regressions without city-level lagged control variables for two reasons. First, although we use lagged city characteristics, there may still be serial correlation with current period values, potentially leading to bad control issues. Second, this can also serve as a balance check. If dropping controls does not significantly change the point estimates, it suggests that the likelihood of omitted variable bias (in this case, location-period level unobserved variables) is low. Tables A4 and A5 demonstrate that the resulting estimates are similar to those in the regressions with control variables. The point estimates remain virtually unchanged. This implies that adding city characteristics does not affect the regression results, further validating the assumption that cities at the border have similar trends.

Table A4: Robustness: TFP Regressions without City-level Controls (OP)

	(1) Local Linear	(2) Poly RD (Poly=1)
Post2003×East	-0.0831**	-0.0701*
	(0.0355)	(0.0425)
City Lagged Controls	N	N
Border FE	Y	Y
Year FE	Y	Y
Firm FE	Y	Y
Observations	131,250	100,054
R-squared	0.1157	0.1116

Notes: The dependent variable is firm-level TFP measured by the Olley and Pakes (1992) method. The regression specifications are identical to Table 2, except we drop all city-level lagged controls. The standard errors are clustered at the firm level. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

Table A5: Robustness: TFP Regressions without City-level Controls (LP)

	(1) Local Linear	(2) Poly RD (Poly=1)
Post2003×East	-0.0506	-0.0867**
	(0.0478)	(0.0438)
City Lagged Controls	N	N
Border FE	Y	Y
Year FE	Y	Y
Firm FE	Y	Y
Observations	85,748	100,054
R-squared	0.1374	0.1446

Notes: The dependent variable is firm-level TFP measured by the Levinsohn and Petrin (2003) method. The regression specifications are identical to Table 2, except we drop all city-level lagged controls. The standard errors are clustered at the firm level. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

A.4 Keeping Slopes Unchanged at the Boundary

Fourth, we change the regression specification to be more parsimonious by keeping slopes unchanged at the boundary. That is, we drop the fourth and the seventh terms in the main regression to have:

$$\begin{aligned} ln(y_{ibct}) = \alpha + \beta_1 East_{ibt} + \beta_2 f(Dist_{ibt}) + Post2003 \times [\delta_1 East_{ibt} + \delta_2 f(Dist_{ibt})] \\ + \beta_4 X_{ct-1} + \phi_{bt} + \gamma_t + \psi_i + \epsilon_{ibct} \end{aligned} \tag{4}$$

Table A6 shows that the results are not changed in this setting. Thus, our conclusion is not sensitive to the choice of the regression discontinuity functional form.

Table A6: Robustness: RD-DID Results with No Slope Change (OP)

	(1) Local Linear	(2) Poly RD (Poly=1)
Post2003×East	-0.0858**	-0.0761*
	(0.0345)	(0.0416)
City Lagged Controls	Y	Y
Border FE	Y	Y
Year FE	Y	Y
Firm FE	Y	Y
Observations	131,250	100,054
R-squared	0.1202	0.1161

Notes: We keep the slopes unchanged around the boundary in this setting. The dependent variable is firm-level TFP measured by the Olley and Pakes (1992) method. The set of lagged city-level control variables includes the log of GDP, the log of population, the log of city area, and the scale of the service sector. The sample in the local linear regression specification is restricted to be within an optimal bandwidth using a constant kernel. The sample in the polynomial RD case is restricted to be within a bandwidth of 40 km around the raw boundary. The standard errors are clustered at the firm level. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

A.5 Thick Border

Fifth, following Michalopoulos and Papaioannou (2014)'s recommendation, we use a thick border in our regression analysis. Provincial borders are often formed by geographical features such as mountains or rivers, and firms at these boundaries may differ significantly from other firms. To address this, we exclude firms within 10 km on both sides of the original provincial borders and extend our bandwidth by 10 km on the far side of the border to preserve total size. This approach mitigates the potential impact of these unique geographic characteristics on our results. Table A7 presents the results using a thick border, and there are no significant changes compared with our baseline setting.

Table A7: Robustness: RD-DID Results with Thick Border (OP)

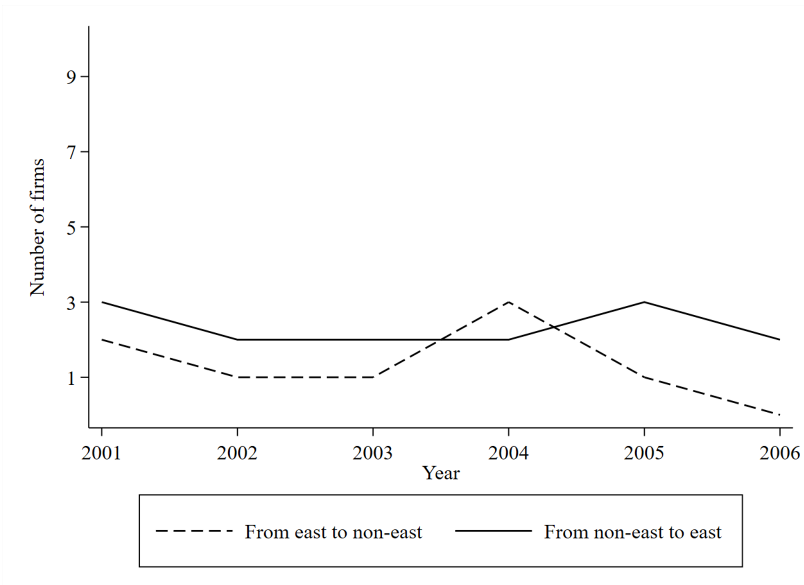
	(1) Local Linear	(2) Poly RD (Poly=1)
Post2003×East	-0.0960	-0.0961*
	(0.0708)	(0.0509)
City Lagged Controls	Y	Y
Border FE	Y	Y
Year FE	Y	Y
Firm FE	Y	Y
Observations	79,668	111,595
R-squared	0.1076	0.1165

Notes: We drop all firms within 10 km of the boundary and create a thick border. The dependent variable is firm-level TFP measured by the Olley and Pakes (1992) method. The set of lagged city-level control variables includes the log of GDP, the log of population, the log of city area, and the scale of the service sector. The sample in the local linear regression specification is restricted to be within an optimal bandwidth using a constant kernel. The sample in the polynomial RD case is restricted to be within a bandwidth of 40 km. The standard errors are clustered at the firm level. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

A.6 Moving Firms

Sixth, our empirical analysis is based on the National Industrial Enterprise Database, a panel dataset that tracks firm movements during the survey years. However, a potential concern is that these relocation decisions may not be exogenous and could be influenced by the inland-favoring land policy. For instance, firms on the eastern side of the border may move to the other side of the boundary to take advantage of cheaper land. If the policy's effect on the local productivity gap is solely a result of this relocation, it may not have a meaningful impact on the economy as a whole.

Figure A1: **Number of Movers from 2001 to 2007**



Notes: This figure shows the number of firms relocating from eastern to inland regions and from inland to eastern regions in each year between 2001 and 2007.

Figure A1 illustrates the yearly count of companies relocating from eastern to inland regions and vice versa between 2001 and 2007 in our dataset. Generally, the number of relocating firms is minimal. For instance, only 3 out of 10,000 firms in our data moved from the east to inland in 2004. Additionally, we do not find any sudden change around the policy year 2003. Table A8 shows the main regression results when we drop all movers. There is no significant change.

Table A8: Robustness: RD-DID Results without Movers (OP)

	(1) Local Linear	(2) Poly RD (Poly=1)
Post2003×East	-0.0827**	-0.0754*
	(0.0356)	(0.0427)
City Lagged Controls	Y	Y
Border FE	Y	Y
Year FE	Y	Y
Firm FE	Y	Y
Observations	131,749	99,953
R-squared	0.1198	0.1161

Notes: We drop all firms changing location. The dependent variable is firm-level TFP measured by the Olley and Pakes (1992) method. The set of lagged city-level control variables includes the log of GDP, the log of population, the log of city area, and the scale of the service sector. The sample in the local linear regression specification is restricted to be within an optimal bandwidth using a constant kernel. The sample in the polynomial RD case is restricted to be within a bandwidth of 40 km. The standard errors are clustered at the firm level. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

A.7 Placebo Test

In this section, we address the spatial spillover issue using two placebo tests. In the first placebo test, we move the boundary west and east to create alternate imaginary boundaries. Then we compare firms on opposite sides of these imaginary boundaries using the main regression. If there are obvious spatial spillovers, that is, if inland firms near the border were also negatively impacted by the policy, we should detect negative policy effects when we move the imaginary boundary to the west. Table A9 shows no evidence for this. In the second placebo test, we redefine the treatment and control groups. We drop all eastern firms and compare only within inland firms. In column (1) of Table A10, we consider inland firms within 5 km of the boundary as the treatment group and inland firms within 5 to 50 km of the boundary as the control group. In column (2) of Table A10, we consider inland firms within 50 km of the boundary as the treatment group and inland firms within 50 to 100 km of the boundary as the control group. Essentially, we are taking inland firms located closer to the boundary as the treatment group and those located further from the boundary as the control group. We find that there is no gaps between these groups that are statistically significant.

Table A9: Robustness: Moving Boundary Placebo Test (OP)

	(1) West 50km	(2) West 100km	(3) East 50km	(4) East 100km
Post2003×East	-0.0204	-0.0060	-0.0215	0.0139
	(0.0647)	(0.0316)	(0.0186)	(0.0142)
City Lagged Controls	Y	Y	Y	Y
Border FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
Firm FE	Y	Y	Y	Y
Observations	51,068	67,420	192,250	272,117
R-squared	0.7411	0.7363	0.7153	0.6968

Notes: The dependent variable is firm-level TFP measured by the Levinsohn and Petrin (2003) method. In columns (1) and (2), we move the boundary to the west by 50 and 100 kilometers, respectively. In columns (3) and (4), we move the boundary to the east by 50 and 100 kilometers, respectively. We use a linear fit as the smoothing function, and the bandwidth is 40 km. The standard errors are clustered at the firm level. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

Table A10: Robustness: Moving Boundary Placebo Test II (OP)

	(1) 0-5km vs 5-50km	(2) 0-50km vs 50-100km
Post2003×East	0.1224	-0.1355
	(0.0748)	(0.0836)
City Lagged Controls	Y	Y
Border FE	Y	Y
Year FE	Y	Y
Firm FE	Y	Y
Observations	39,706	69,068
R-squared	0.7667	0.7603

Notes: The dependent variable is firm-level TFP measured by the Olley and Pakes (1992) method. The set of lagged city-level control variables includes the log of GDP, the log of population, the log of city area, and the scale of the service sector. The sample in the local linear regression specification is restricted to be within an optimal bandwidth using a constant kernel. In column (1) we compare inland firms within 5 km of the boundary with inland firms within 5 to 50 km of the boundary. In column (2) we compare inland firms within 50 km of the boundary with inland firms within 50 to 100 km of the boundary. The standard errors are clustered at the firm level. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

A.8 Robustness Checks for The WTO Effect

Seventh, China joined the WTO at the end of 2001, leading to significant changes in the country's economic structure. Despite occurring two years before the inland-favoring land supply policy, we remain concerned about potential confounding effects from the reduction in trade barriers, which may have influenced eastern and inland firms differently. To address this issue, we conduct the TFP regression using only firms with zero exports, as they should be the least affected by any WTO effects. Additionally, we run the main regression while controlling for firm-level exporting to eliminate any WTO-related influence.

Table A11: Robustness: TFP Regressions without Exporting Firms (OP)

	(1) Local Linear	(2) Poly RD (Poly=1)
Post2003×East	-0.0894**	-0.1079**
	(0.0406)	(0.0487)
City Lagged Controls	Y	Y
Border FE	Y	Y
Year FE	Y	Y
Firm FE	Y	Y
Observations	105,161	79,951
R-squared	0.1229	0.1204

Notes: The dependent variable is firm-level TFP measured by the Olley and Pakes (1992) method. The regression specifications are identical to Table 2. We drop all firms with positive exports. The standard errors are clustered at the firm level. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

Table A12: Robustness: TFP Regressions without Exporting Firms (LP)

	(1) Local Linear	(2) Poly RD (Poly=1)
Post2003×East	-0.1169**	-0.1395***
	(0.0550)	(0.0502)
City Lagged Controls	Y	Y
Border FE	Y	Y
Year FE	Y	Y
Firm FE	Y	Y
Observations	68,439	79,951
R-squared	0.1453	0.1533

Notes: The dependent variable is firm-level TFP measured by the Levinsohn and Petrin (2003) method. The regression specifications are identical to Table 2. We drop all firms with positive exports. The standard errors are clustered at the firm level. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

The regression results are displayed in Tables A11, A12, A13, and A14. Our main conclusions remain consistent. We also find that a firm's exporting activity positively relates to its productivity, which aligns with predictions in the trade literature (Bernard et al., 2007, 2018).

Table A13: Robustness: TFP Regressions Controlling for Exporting (OP)

	(1) Local Linear	(2) Poly RD (Poly=1)
Post2003×East	-0.0719**	-0.0661
	(0.0355)	(0.0425)
log(Export)	0.0157***	0.0160***
	(0.0013)	(0.0015)
City Lagged Controls	Y	Y
Border FE	Y	Y
Year FE	Y	Y
Firm FE	Y	Y
Observations	131,250	100,054
R-squared	0.1221	0.1181

Notes: We additionally control for firm-level exports in this regression. The dependent variable is firm-level TFP measured by the Olley and Pakes (1992) method. The regression specifications are otherwise identical to Table 2. The standard errors are clustered at the firm level. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

Table A14: Robustness: TFP Regressions Controlling for Exporting (LP)

	(1) Local Linear	(2) Poly RD (Poly=1)
Post2003×East	-0.0383	-0.0760*
	(0.0477)	(0.0436)
log(Export)	0.0253***	0.0256***
	(0.0016)	(0.0015)
City Lagged Controls	Y	Y
Border FE	Y	Y
Year FE	Y	Y
Firm FE	Y	Y
Observations	85,748	100,054
R-squared	0.1465	0.1542

Notes: We additionally control for firm-level exports in this regression. The dependent variable is firm-level TFP measured by the Levinsohn and Petrin (2003) method. The regression specifications are otherwise identical to Table 2. The standard errors are clustered at the firm level. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

A.9 Robustness Checks for Subsidy and Tax Policies

Eighth, we attempt to rule out the effects of other concurrent subsidy and tax policies that may have been implemented alongside the land reform. Apart from the land supply policy, the Chinese government also enacted other inland-favoring measures to promote inland economic growth, such as manufacturing subsidies. We conduct the primary regression using firm-level government subsidies as the outcome variable to check whether relative subsidies changed for firms at the border during the same year the inland-favoring land policy was introduced. Table A15 indicates that firms on either side of the border received similar government subsidies before and after 2003. We then carry out the firm-level TFP regressions with additional controls, including city-level central government subsidies per capita, firm subsidies from the government, and firm-level taxes paid to the government. Tables A16 and A17 demonstrate that the main results remain consistent across all regression settings.

Table A15: Robustness: RD-DID Results on Firm-level Subsidies

	(1) Local Linear	(2) Poly RD (Poly=1)
Post2003×East	-0.0034	-0.0015
	(0.0022)	(0.0019)
City Lagged Controls	Y	Y
Border FE	Y	Y
Year FE	Y	Y
Firm FE	Y	Y
Observations	77,332	100,054
R-squared	0.0026	0.0023

Notes: The dependent variable is firm-level subsidies. The set of lagged city-level control variables includes the log of GDP, the log of population, the log of city area, and the value added to the service sector. The sample in the local linear regression specification is restricted to be within an optimal bandwidth using a constant kernel. The sample for the polynomial RD case is restricted to be within a bandwidth of 40 km around the original boundary. The standard errors are clustered at the firm level. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

Table A16: RD-DID Results with Firm-level Subsidy and Tax Controls (OP)

	(1) Local Linear	(2) Poly RD (Poly=1)
Post2003×East	-0.0728**	-0.0652
	(0.0339)	(0.0404)
Tax	-1.7408***	-1.7556***
	(0.0213)	(0.0244)
Subsidy	-0.9215***	-0.9819***
	(0.0933)	(0.1061)
City Lagged Controls	Y	Y
Border FE	Y	Y
Year FE	Y	Y
Firm FE	Y	Y
Observations	131,250	100,054
R-squared	0.2541	0.2534

Notes: The dependent variable is firm-level TFP measured by the Olley and Pakes (1992) method. We additionally control for firm-level subsidies and firm-level taxes in these regressions. The regression specifications are identical to Table 2. We drop city-level lagged controls. The standard errors are clustered at the firm level. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

Table A17: RD-DID Results with Firm-level Subsidy and Tax Controls (LP)

	(1) Local Linear	(2) Poly RD (Poly=1)
Post2003×East	-0.0413	-0.0810*
	(0.0454)	(0.0416)
Tax	-1.7360***	-1.7640***
	(0.0274)	(0.0250)
Subsidy	-0.9342***	-0.9218***
	(0.1242)	(0.1110)
City Lagged Controls	Y	Y
Border FE	Y	Y
Year FE	Y	Y
Firm FE	Y	Y
Observations	85,748	100,054
R-squared	0.2712	0.2819

Notes: The dependent variable is firm-level TFP measured by the Levinsohn and Petrin (2003) method. We additionally control for firm-level subsidies and firm-level taxes in these regressions. The regression specifications are identical to Table 2. We drop city-level lagged controls. The standard errors are clustered at the firm level. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

A.10 Clustering Standard Errors

In the main regression, we cluster the standard errors at firm-level, which is the cross-sectional unit of our panel data. This is recommended by Angrist and Pischke (2009); Abadie et al. (2023). A potential concern is that firms located close to each other may be exposed to common shocks, which can result in spatial correlation of error terms. To capture this correlation, we cluster the standard error at province-level. We run this regression after dropping all firms which changed location during the sampling period (movers) due to technical issues. Table A18 shows that we still have significant (or marginally significant) estimates. These estimates correspond to the estimates in Table A8. We claim that this is a very conservative estimation of our standard error because we have a dataset of all above-scale enterprises in China. If we do not consider our non-negligible sample size compared with the population, the standard error is likely overestimated (Abadie et al., 2020). Clustering at too high a level is also not recommended by Abadie et al. (2023). Therefore, we cluster the standard error at firm-level in our main regression.

Table A18: RD-DID Results on TFP (Clustering at Province-level)

	(1) Local Linear	(2) Poly RD (Poly=1)
Post2003×East	-0.0827	-0.0754*
	(0.0503)	(0.0404)
City Lagged Controls	Y	Y
Border FE	Y	Y
Year FE	Y	Y
Firm FE	Y	Y
Observations	131,749	99,953
R-squared	0.1198	0.1161

Notes: The dependent variable is firm-level TFP measured by the Olley and Pakes (1992) method. The set of lagged city-level control variables includes the log of GDP, the log of population, the log of city area, and the scale of the service sector. The sample in the local linear regression specification is restricted to be within an optimal bandwidth using a constant kernel. The sample in the polynomial RD cases is restricted to be within a bandwidth of 40 km around the raw boundary. The standard errors are clustered at the province level. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.