PIER

PUEY UNGPHAKORN INSTITUTE
FOR ECONOMIC RESEARCH

Entry and Spatial Competition of Intermediaries: Evidence from Thailand's Rice Market

by

Bunyada Laoprapassorn

January 2022
Discussion Paper
No. 170

Entry and Spatial Competition of Intermediaries: Evidence from Thailand's Rice Market*

Bunyada (Mos) Laoprapassorn
University of Michigan

December 5, 2021

Click here for latest version

Abstract

How does the market power along the agricultural value chains mediate the effects of policies on the welfare of farmers? Using microdata on farmers and rice mills in Thailand, I document heterogeneity in the spatial density of rice mills. I further provide reduced-form evidence that a one standard deviation increase in local competition among rice mills leads to a 7.7% increase in farmer prices. Informed by the empirical findings, I propose and estimate a quantitative spatial model that accounts for the market power and entry-location choices of intermediaries. I then simulate two policy counterfactuals. I find that gains to farmers from a country-wide improvement in road infrastructure are regressive; the percentage increase in income of the top decile farmers is on average 11% larger than that of the bottom decile. Changes in the entry decisions of the rice mills further exacerbate the regressive effect, more than doubling the gap between the percentage change in income of the top and bottom decile farmers. The second counterfactual simulation shows that the market power of intermediaries could lead to a lower than socially optimal level of technology adoption among farmers.

*I am deeply indebted to Lauren Bergquist, Andrei Levchenko, Sebastian Sotelo, and Jagadeesh Sivadasan for their advice, guidance, and encouragement. I am extremely grateful to Dominick Bartelme, Zach Brown, Ying Fan, and Hoyt Bleakley for their advice and insightful comments. I would like to thank Mariá Aristizábal-Ramírez, Barthélemy Bonadio, Jaedo Choi, Michelle Lam, Nafisa Lohawala, Nishaad Rao, and seminar participants at the University of Michigan for helpful comments and suggestions. I am grateful for the research support from the Office of Agricultural Economics, the Bureau of Trade and Economic Indices, the Department of Business Development, the National Statistical Office, and Puey Ungphakorn Institute for Economic Research. E-mail: bunyada@umich.edu.

1 Introduction

How does the market power along the agricultural value chains mediate the effects of policies on the welfare of farmers? A large proportion of the population in developing countries relies on farming as a source of livelihoods, with agriculture accounting for 60% of employment in low-income countries and 29% in middle-income countries (World Bank, 2021). One of the prominent features of agriculture in developing countries is a large number of small farmers and a small number of intermediaries, suggesting the possible existence of market power along the value chain. This is particularly true for many staple crops that need to be processed before they become edible, such as rice and wheat. The fixed cost and economies of scale arising from the intermediate processing activities give rise to large intermediaries; for instance, in the rice market, rice mills emerge. Local market concentration around these large-scale intermediaries contrasts the fragmented nature of small-scale farmers. This can potentially allow intermediaries to enjoy significant market power and suppress the prices that farmers receive. Furthermore, these intermediaries make strategic entry and location choices, which are influenced by the heterogeneous local market sizes and transport costs, leading to heterogeneity in the spatial density of intermediaries across the country. This, in turn, results in variations in the level of local competition of intermediaries and possibly variations in the prices that farmers receive.

Assessing how policies impact farmer income requires understanding both how intermediaries' entry decisions change in response to those policies and how subsequent changes in the spatial density of intermediaries impact prices that farmers receive. For instance, consider an improvement in road infrastructure, which reduces trade costs. For a given spatial density of intermediaries, we would expect lower trade costs to increase prices that farmers receive. However, our conjecture about the overall effect of this policy becomes ambiguous once we account for intermediaries' entry responses. On the one hand, because upward pressure on farmer prices reduces the profitability of the intermediaries, intermediaries may respond by exiting the market, leading to lower competition and lower farmer prices. On the other hand, higher farmer prices could incentivize farmers to increase their output. A higher quantity of rice traded means higher total profit for the intermediaries, which encourages more intermediaries to enter, increasing the spatial competition. Therefore, to understand the effects of shocks on the welfare of farmers, it is necessary to take into account both the intermediaries' entry-location decisions and the subsequent impact on their market power.

In this paper, I develop a framework to study the consequences of policies, such as an improvement in road infrastructure, on the welfare of farmers. I approach the question in the context of the rice market in Thailand, focusing on the market power of rice mills and its impact on farmers. I assemble a micro-level dataset on locations of farmers and rice mills, prices, and rice production. In the first part of the paper, I begin by providing two motivating observations of the entry, exit, and spatial distribution of rice mills. First, I document the heterogeneity in the spatial distribution of rice mills; productive areas tend to have a higher density of mills on average, signifying the existence of economies of scale for rice mills and heterogeneity in the spatial competition among rice mills. Second, I show that there is an active margin of entry and exit of rice mills, which leads to changes in the density of rice mills and their spatial competition.

I then provide reduced-form, causal evidence that spatial competition among rice mills affects farmer prices. Similarly to Macchiavello and Morjaria (2021), I instrument for the local competition among rice mills using the productivity of the farmers in the neighboring area. Conditional on the farmer's own productivity, a farmer surrounded by more productive neighbors is surrounded by a larger number of mills and experiences a higher degree of competition among the mills (first stage). I establish that a one standard deviation increase in local competition results in a 7.7% increase in farmer prices[1] (second stage).

Next, I develop a quantitative structural model grounded in these empirical findings. I adopt a framework that captures both the uneven spatial distribution of rice mills and the effect of spatial competition among intermediaries on farmer prices. In the model, rice farmers grow rice on plots with heterogeneous productivity. Post-harvest, farmers optimally choose mills to sell the rice they have grown. The mills then sell rice to retailers at fixed and exogenous retail prices. Following Chatterjee (2020), I model the spatial market power of rice mills using a Nash bargaining framework, in which spatial market power arises because of the transport cost and the physical distances between rice mills. Farmer price is determined through Nash bargaining between the farmer and the rice mill. If the mill and the farmer cannot reach an agreement, the farmer transports and sells rice to another mill. Since transport is subject to iceberg trade cost, the threat point of the farmer in the Nash bargaining process is the value they would get if they sell rice to another mill after accounting for the transport cost. Larger distances between mills mean that farmers have lower threat points, resulting in lower equilibrium farmer prices. Therefore, mills situated in areas with a higher mill density have lower market power and pay higher prices to farmers.

Heterogeneity in the spatial distribution of rice mills arises because rice mills make strategic entry and location decisions that are influenced by the heterogeneity in market sizes and remoteness of the locations. To endogenize the entry decisions of mills, I employ a static entry game with incomplete information about the competitors' entry cost as in Seim (2006). While making entry-location decisions, mills face a trade-off between access to farmers and the level of competition. On the one hand, mills want to locate in places with higher quantity of rice in order to make higher profits. On the other hand, mills want to avoid locating close to other mills so that they face lower competition. In equilibrium, areas with higher aggregate rice output, which I term productive areas, can support a higher number of mills, leading to higher mill density. As a result, mills in productive areas face a higher level of local competition, and farmers in productive areas receive higher prices relative to

[1] I define farmer price to be the price that farmers receive from a mill. Farmer price is different from farm gate price, which is the farmer price subtracted by the cost of transporting rice from the farm to the mill.

those in unproductive areas.

Due to the interdependencies in the mills' decisions, solving the static entry game for an exact equilibrium is a computationally intractable problem. A mill's profit in a location depends on both the number of mills in that location and the number of mills in other locations. Therefore, while making an entry-location decision, mills have to consider all the potential location choices of all other potential entrants. Since there are multiple locations and multiple potential entrants, the space of mills' actions becomes too large to be computationally feasible even when the number of locations and the number of potential entrants are not too large. To address this dimensionality challenge, I obtain an approximation of an equilibrium of the static entry game using the method proposed by Aguirregabiria and Vicentini (2016). In particular, I assume that the location-specific variable profit function is a second order polynomial in the number of entrants in each location. For a game with 180 locations and 2,000 potential entrants in this paper, this approach reduces the dimensionality down from over 2×10^{268} to under 7×10^4.

I quantify the model using micro-level data on locations of mills and farmers, prices, and rice production. I solve the model sequentially and estimate parameters in the Nash-bargaining problem and the entry game separately using a two-step procedure. In the Nash bargaining problem, the structural relationship between farmer prices, density of rice mills, and retail prices depends on the bargaining power of farmers and the transport costs. To estimate parameters governing this relationship, I first take the number and the locations of rice mills in the data as given, which pins down the distances between the rice mills. With the observed mill density in the data, I estimate parameters that determine farmer prices in the Nash bargaining problem using simulated method of moments. Specifically, I match coefficients of an auxiliary regression of farmer prices on the distance to the mill, the retail price, and the maximum alternative price within a 100 kilometer radius, as well as the parameters of the distribution of farmer prices in the data. With these estimated parameters in the Nash-bargaining problem, I then estimate the parameters in the entry problem to match the spatial distribution of rice mills observed in the data. Because the mill distribution is an equilibrium in the entry game, I estimate the parameters using the nested fixed point maximum likelihood algorithm. Specifically, for a given set of parameters in the entry problem, I solve for a fixed point in the entry game and nest the fixed-point algorithm into the maximum likelihood routine.

I then use the estimated model to conduct two counterfactual experiments, which highlight how the intermediaries' entry decisions and their subsequent market power impact farmers' income. First, motivated by the strategic plan of the Department of Highways of Thailand for investment in road infrastructure in Thailand, I simulate a country-wide reduction in trade costs. Second, motivated by ongoing projects in Thailand that promote new farming practices, I study how intermediaries impact farmers' decisions to adopt new technology. For each counterfactual scenario, I conduct two exercises. In the first exercise, I examine the effect on farmers' income while holding the number of mills in each location the same as the baseline scenario. Then, I study the change in farmers' income after allowing mills to change their entry-location decisions.

In the first counterfactual scenario, based on the Department of Highways' 2017 strategic plan, I simulate a 9.09% country-wide reduction in the iceberg trade costs. I find that the intermediaries' strategic entry decisions and spatial market power have significant distributional consequences. While the aggregate farmer income increases by 16.75%, the percentage increase in income of the top decile farmers is on average 25% larger than the percentage increase in income of the bottom decile. Productive locations attract a larger number of intermediaries, leading to lower market power of intermediaries. Therefore, a larger portion of the gains is passed from intermediaries to farmers in productive areas relative to farmers in unproductive areas.

Furthermore, changes in mills' entry-location decisions in response to the policy are regressive. A reduction in trade costs can affect mills' entry decisions in several ways. First, lower trade costs reduce the costs of accessing nearby mills and increase the threat points for farmers in the Nash bargaining, which force intermediaries to offer higher prices, reducing their per-unit profit. Second, lower trade costs increase the quantity of rice that farmers produce and trade to mills, increasing the total profit of rice mills. In unproductive locations, the increase in farmers' output is not sufficient to offset the effect of lower profit margin, leading to a net decline in profit. In contrast, in productive locations, the additional profit brought about by higher output exceeds the loss from a lower profit margin. Therefore, productive locations become relatively more attractive; the number of mills rises in productive locations and falls in unproductive locations. Consequently, the percentage increase in farmer prices in productive areas becomes larger than what would have been if there had been no entry response of intermediaries and vice versa. As a result, farmers in productive areas enjoy a larger percentage increase in income relative to unproductive farmers. Ignoring the entry response of intermediaries leads us to underestimate the gap between the percentage increase in income of the top decile and the bottom decile farmers by 53%.

In the second counterfactual analysis, I consider how intermediaries' entry decisions and spatial market power impact farmers' adoption of new technology. Based on an ongoing project in Thailand, I model farmers' decisions of whether to adopt a new agricultural technology that will increase their productivity by 30% with an upfront investment cost. I illustrate the existence of multiple equilibria, one of which results in a level of investment that is lower than socially optimal. Multiple equilibria arise because farmers' return to investment depends on the prices they receive, which in turn depends on the market power of the intermediaries. Because farmers are small relative to the rice market, they cannot individually directly influence the spatial distribution of rice mills. Likewise, farmers do not internalize how their individual investment decisions could collectively lead to a higher number of mills, which would increase farmer prices and return to investment. I demonstrate that if farmers believe that no other farmers will invest in the new technology, their estimated return will be sufficiently low that they believe it is not worthwhile to invest. On the other hand, if farmers collectively invest in the new technology, more intermediaries enter, driving up the return such that it would be worthwhile for 62% of the eligible farmers to invest. Because farmers do not internalize the strategic complementarity between their investment decisions and the intermediaries' entry decisions, farmers can end up in an equilibrium with a lower than socially optimal level of investment. Thus, subsidies for technology adoption could be welfare improving.

This paper contributes to the existing literature in several dimensions. The first is a growing body of work that examines the market power of intermediaries in agriculture. Tomar (2016) and Chatterjee (2020) estimate sizable welfare gains for the farmer as a result of policy reforms that reduce the market power of intermediaries. Nevertheless, experimental evidence on intermediary market structure shows mixed results; while results in Casaburi and Reed (2021) suggest a competitive intermediary sector, Bergquist and Dinerstein (2020) estimate significant market power of intermediaries. The majority of the existing literature focuses on measuring the market power among intermediaries within a particular marketplace without accounting for the spatial locations of the marketplace. While Chatterjee (2020) measures the competitiveness across marketplaces, in his paper, the locations of intermediaries are exogenously determined by the government and fixed across time. This paper contributes to the literature in two aspects. Empirically, I provide reduced-form, causal evidence from the universe of intermediaries in a market with active entry and exit, showing that the interaction of economic geography and spatial competition among intermediaries result in variations in the local market power of intermediaries. When it comes to theory, instead of taking the intermediaries' locations as exogenously given, this paper endogenizes intermediaries' entry and location decisions. As I illustrate in my motivating observations, entry decisions of intermediaries could potentially be important as intermediaries are spatially unevenly distributed. Modeling the entry-location decisions allows me to examine how the extensive margin of intermediaries changes in response to external shocks and their distributional implications on farmers' income.

This paper also contributes to a growing literature in international trade that analyzes interdependencies in firm-level decisions. While many papers have studied the interdependencies in firm's entry decisions arising from granularity of firms (Eaton et al., 2012; Gaubert and Itskhoki, 2021; Gaubert et al., 2021), existing work limits the interactions across firms to be *within* a market. In such a case, a firm's decision to enter a market can be analyzed separately for each market by assuming sequential entry. Since there are no interdependencies in firms' decisions across markets, the problem remains computationally feasible. The other strand of spatial and trade literature has considered the interdependencies in sourcing decisions across suppliers *within* a firm. By focusing on interactions within a firm, Antras et al. (2017) are able to limit their analysis to instances where importing decision of a firm exhibits strategic complementarities across suppliers, which means that the addition of a supplier to a firm's sourcing strategy increases the marginal gain from adding another supplier. The complementarities in sourcing decisions of a firm allow the problem to be solved using an algorithm developed by Jia (2008). However, the approach cannot be extended to instances where firms' decisions are strategic substitutes. Hoang (2020) estimates the fixed costs and sunk costs of sourcing inputs using moment inequalities, which can be applied both when firms' decisions are strategic complements and strategic substitutes, but does not conduct any counterfactual analysis. This paper studies the location decisions of the rice mills, taking into account the strategic interactions both *across* mills and *across* locations. Interdependencies across locations mean that I cannot analyze the mills' entry decisions separately by locations. Likewise, interdependencies across mills mean that intermediaries' entry strategies are strategic substitutes. I overcome the combinatorial challenge in mills' entry decisions by adopting the approach proposed by Aguirregabiria and Vicentini (2016).

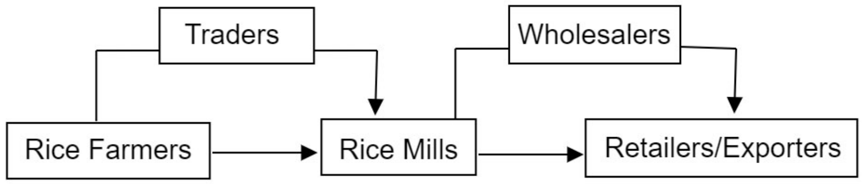
Additionally, this paper contributes to the literature that analyzes the distribution of gains from increased integration under variable markups. The majority of the existing literature focuses on the variable markups brought about by the producers (Badinger, 2007; Edmond et al., 2015; Feenstra and Weinstein, 2017; Melitz and Ottaviano, 2008). While Chatterjee (2020) estimates gains from increased integration arising from changes in the spatial market power of intermediaries, the setting of the paper is in India, where locations of intermediaries are exogenously constrained by state licensing and where there are no entry responses of intermediaries. A novelty in this paper is that I examine how the strategic entry-location decisions and spatial variations in monopsony power of intermediaries impact the distribution of gains from increased integration. I show that if there is sufficiently large heterogeneity in underlying geographical features across space, strategic entry decisions of intermediaries will result in the varying spatial market power of intermediaries across space, leading to a regressive distribution of gains from increased integration.

The rest of the paper proceeds as follows. Section 2 describes the structure of the rice market in Thailand and the data sources. Section 3 presents motivating observations about the rice mills in Thailand and reduced-form evidence of the spatial market power of intermediaries that motivate my choice of model. In section 4, I develop a quantitative spatial model informed by the empirical findings. Section 5 describes how I recover structural parameters using the two-step estimation strategy. Section 6 presents findings from the two counterfactuals based on policies in Thailand. Section 7 concludes.

2 Background and Data

Rice is a staple crop in Thailand; 46% of the cropland in Thailand is devoted to rice production (Office of Agricultural Economics, 2019). Figure 1 shows the structure of the rice market in Thailand. Within the rice value chain, rice mills are the key players in the midstream-level activities. In this paper, I focus on the relationship between the farmers and the rice mills. The term intermediaries in this paper refers to rice mills. In Thailand, there are very few regulatory barriers that prevent businesses from entering the rice market. Therefore, rice mills are free to make their entry and location decisions based on the profitability and the entry cost.

Figure 1: Rice Market Structure



Examining how entry decisions of rice mills affect their spatial market power and the farmers requires micro-level geospatial data on the rice mills and farmers. Since there is no publicly available dataset, I assembled a micro-level dataset on the farmers, the rice mills, the general cropping patterns, and rice prices between 2008-2018 from various sources. I outline below the primary datasets that I use.

Rice Mills: Data on rice mills consists of the annual balance sheets of all firms registered with Thailand Standard Industrial Classification (TSIC) code for "rice milling" that are submitted to the Department of Business Development, Ministry of Commerce. This is an unbalanced panel from 2007-2019. I identify the locations of the mills from their registered addresses using Google Maps API.

Farmers: Farmer-level data on prices that farmers received, the input usage, and the quantity of rice sold are taken from the socio-economic and labor survey on agricultural households between 2008-2018. The survey was conducted annually by the Office of Agricultural Economics. I identify the locations of the farmers at the sub-district level.[2] The latitudes and the longitudes of the sub-districts are obtained from the Department of Provincial Administration and supplemented by Google Maps API.

Rice production and Yields: Data on rice production and yields come from the Office of Agricultural Economics. I have data on the monthly quantity of rice harvested at the province level between 2008-2018. Additionally, I have the annual quantity of rice at the district level between 2012-2018. I supplement this dataset with data from the 2003 and 2013 Agricultural Census and the 2008 and 2018 Agriculture Intercensal Survey provided by the National Statistical Office.

Retail Prices: I obtain data on retail prices from the Bureau of Trade and Economic Indices, Ministry of Commerce. Retail prices are collected on a monthly frequency and are available at the province level.

In addition, I supplement my dataset with high-spatial-resolution data. I obtain data on crop-suitability-index from the Global Agro-Ecological Zone (GAEZ) v.3 and data on area equipped

[2]The administrative levels in Thailand are provinces, districts, and sub-districts. There are 77 provinces, 928 districts, and 7,435 sub-districts (Department of Enviromental Quality Promotion, n.d.).

with irrigation from AQUAMAPS. I further utilize the terrain ruggedness measure developed by Nunn and Puga (2012). Lastly, I obtain data on the population density from the Gridded Population of the World (GPWv4) provided by NASA Socioeconomic Data and Applications Center (SEDAC).

3 Characteristics of Thailand's Rice Market

In this section, I present key characteristics of rice mills and farmer prices in Thailand that motivate my choice of model. First, I provide two motivating observations about the endogenous entry and location decisions of the rice mills. I show that 1) there is a positive correlation between spatial density of mills and quantity of rice grown in the area, implying heterogeneity in the level of local competition among rice mills and the existence of economies of scale for rice mills, and 2) there is an active margin of entry and exit of rice mills. Having done so, I provide reduced-form, causal evidence that spatial competition among rice mills impacts farmer prices.

3.1 Entry, Exit, and Spatial Density of Rice Mills

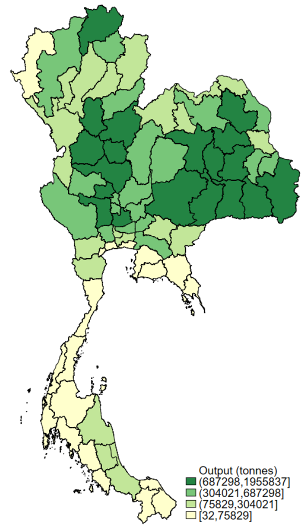
I first present two observations about the rice mills in Thailand that motivate me to endogenize entry decisions of rice mills. First, there is an uneven spatial distribution of rice mills across the country. Figure 2(a) displays the locations of rice mills in 2018, while Figure 2(b) shows the quantity of rice harvested in each province in 2018. We can see a large number of mills concentrated in the central region and the northeastern region. In particular, Figure 2(c) shows a positive correlation between the density of rice mills and the quantity of rice grown, which is indicative of the existence of fixed costs and economies of scale. Although there is no regulatory barrier for rice mills to enter the rice market in Thailand, the presence of fixed costs means that mills would only enter if they can achieve a certain level of sales. Therefore, areas with more productive farmers will experience higher mill densities. Higher spatial density of rice mills in productive areas suggests that mills in productive areas face a higher level of spatial competition relative to those in unproductive areas. Therefore, the observed distribution of rice mills along with its positive correlation with the quantity of rice output are important features that the model should reflect.

Second, there is an active margin of entry and exit of rice mills. Table 1 reports the summary statistics of the rice mills between 2007-2019. While there are on average about 1,011 mills operating each year, the margin of entry and exit fluctuates across the years; the net entry rate varies from -4.7% to 4.7%, resulting in the number of mills across the sample period that ranges between 944 and 1,084. Entry and exit of rice mills mean changes in spatial density of mills, impacting the rice mills' level of spatial competition and market power. Furthermore, there is heterogeneity in the entry and exit of rice mills across the provinces. Figure 3 displays variations in the average net entry rate of each province between 2008-2019, indicating heterogeneity in changes in spatial density of rice mills across the country. Given that changes in spatial density of mills may impact farmer prices, entry and exit form an important margin that the model needs to capture.

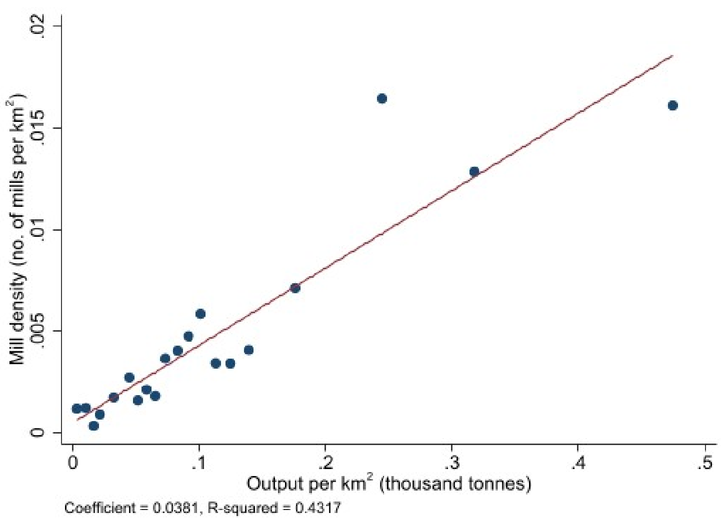
Figure 2: Spatial Distribution of Rice Mills

(a) Locations of Mills in 2018

(b) Output in 2018



(c) Binned Scatter Plot of Mill Density and Average Output per km^2 in Each Province



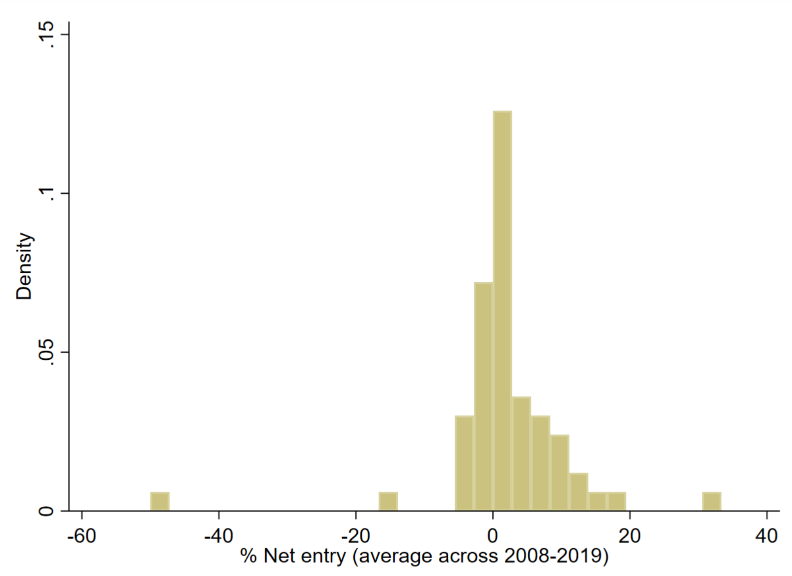
Notes: Panel (a) plots the locations of rice mills in 2018. Panel (b) shows the quantity of rice produced per province in 2018. Panel (c) shows a binned scatter plot of the average number of mills per km^2 in each province against the average output per km^2 (thousand tonnes) in each province between 2008-2018.

Table 1: Summary Statistics of Rice Mills

	Mean	Std. Dev.	Min	Max
No. of mills	1,010.85	45.14	944.00	1,084.00
% entering	8.23	2.33	5.65	13.24
% exiting	8.34	1.41	6.86	10.95
% net entry	-0.06	2.90	-4.65	4.72

Notes: This table reports the average of the total number of mills, entry rate, exit rate and net entry rate in Thailand between 2007-2019.

Figure 3: Histogram of Average % Net Entry in Each Province



Notes: Net entry rate of mills in each province, averaged across 2008-2019.

3.2 Spatial Competition and Farmer Price

Having shown that there are variations in the spatial distribution of rice mills across Thailand, I now provide reduced-form evidence that a higher density of rice mills results in higher prices that farmers receive. In a market with imperfect competition among rice mills, higher mill density means a higher level of local competition, which reduces the mills' market power. Therefore, farmers situated in areas with a higher density of rice mills benefit from higher competition among rice mills.

I first construct a measure that captures the degree of spatial competition among rice mills. I define local competition as the sum of the number of rice mills surrounding a farmer weighted by the inverse of the distances between the farmer and the mills. The local competition among rice mills that farmer f experiences is the sum of all the mills within 100 km radius from the farmer, where the weights are inverse to the distance between the farmer and the mills. Specifically,

$$COMP_{ft} = \sum_{m \in \mathcal{M}_{100\text{km},t}} \left\{ \frac{1}{distance_{fm}} \right\} , \tag{1}$$

where $\mathcal{M}_{100\text{km},t}$ is the set of all the mills situated within 100 km radius around farmer f at time t and $distance_{fm}$ is the geodesic distance between farmer f and mill m. A farmer situated close to a large number of rice mills will exhibit a higher competition measure. This measure is similar to the competition measure in Chatterjee (2020) and the market access measure in Donaldson and Hornbeck (2016).[3] As a robustness check, I also use an alternative measure of competition, defining local competition as simply the number of mills within a 100 km radius from the farmer.

Having constructed the competition measure, I can show the relationship between local competition and farmer prices using the following reduced-form specification:

$$p^f_{frct} = \beta_0 + \beta_1 COMP_{ft} + \beta_2' Z_{ft} + \gamma_r + \gamma_c + \gamma_t + \varepsilon_{frct} \ , \tag{2}$$

where p^f_{frct} is the log of price that farmer f in region r receives for crop type c at time t. Z_{ft} denotes other control variables. γ_r and γ_c controls for the region and crop-type[4] fixed effects, and γ_t controls for the time fixed effects. To allow spatial correlation, I adjust the standard errors as in Conley (1999).

Table 2 reports the results from the OLS regressions.[5] Column 1 reports the results when I only control for the fixed effects, namely the year fixed effects, the fixed effects for the month in which the farmers sell the majority of their crops, the region fixed effects, and the crop-type fixed effects. There is a positive and significant relationship between local competition and farmer prices; a one standard deviation increase in competition among rice mills corresponds to a 1.4% increase in price that farmers receive. To account for other factors that may influence farmer prices, I add in four more control variables. Since the local supply of rice could affect local price, I control for the quantity of rice that the farmer sells and the quantity of rice harvested in the province. I also control for land suitability for growing rice using the crop-suitability-index from the FAO GAEZ dataset. Likewise, I control for the distance between the farmer and Bangkok, which is the capital of Thailand and close to the largest port in Thailand. The results after controlling for other variables

[3] Chatterjee (2020) creates a competition measure for a marketplace as the weighted sum of other marketplaces within the same state since the Agriculture Produce and Marketing Committee Acts of Indian states prohibit farmers from selling their output to government-regulated marketplaces outside of their own state. There is no regulatory restriction on whom the farmers could sell the rice to in Thailand, though in practice, farmers generally sell rice to nearby mills. Therefore, I use 100 km as a cutoff for the competition measure. As robustness checks, I conduct the same analysis using an analogous competition measure with different distance cutoffs. Results can be found in Appendix A.1.

[4] The crop types I am able to observe in the data are in-season white rice, out-of-season white rice, in-season sticky rice, and out-of-season sticky rice.

[5] The surveys in 2016-2018 do not contain data on the month in which farmers sell the majority of their rice. Therefore, I only use farmer prices in 2008-2015 in my regressions.

are shown in column 2. After controlling for other factors, a one standard deviation increase in local competition corresponds to a 2.3 % increase in farmer prices.

Table 2: OLS Results

	log(farmer price)			
	(1)	(2)	(3)	(4)
COMP (std)	0.014**	0.023***		
	(0.006)	(0.008)		
No. of mills (std)			0.036**	0.036**
			(0.014)	(0.015)
log(crop sold)		-0.006	-0.006	-0.006
		(0.005)	(0.005)	(0.005)
log(province-level output)		-0.012	-0.013	-0.012
		(0.008)	(0.008)	(0.008)
Crop-suitability-index		0.010*	0.008	0.009
		(0.006)	(0.006)	(0.006)
log(distance to BKK)		0.024	0.040*	0.040*
		(0.017)	(0.023)	(0.023)
log(distance to nearest mill)				0.003
				(0.003)
Year FE	Yes	Yes	Yes	Yes
Month-of-highest-sales FE	Yes	Yes	Yes	Yes
Region FE	Yes	Yes	Yes	Yes
Crop-type FE	Yes	Yes	Yes	Yes
N	54,323	54,252	54,252	54,252
R^2	0.509	0.511	0.512	0.512

Notes: This table reports the OLS estimates of equation (2). COMP (std) is the standardized competition measure constructed using equation (1). No. of mills (std) is the standardized number of mills within 100 km from the farmer. All regressions use data from 2008 to 2015. Standard errors are adjusted to allow for spatial clustering as in Conley (1999), with a bandwith of 1.5 degree using Bartlett kernel. $*p < 0.1$, $**\ p < 0.05$, $***\ p < 0.01$

As robustness checks, I run the same regression using an alternative measure of local competition. Instead of using the measure constructed using equation (1), I use the number of rice mills within a 100 km radius from the farmer. Column 3 of Table 2 reports results when I use an alternative measure of competition. Column 3 shows that the positive and significant relationship between competition and farmer prices is not sensitive to the competition measure I use. I perform additional robustness checks using the average farmer prices betweeen 2016-2018 as an additional control variable and using alternative competition measures with different cutoffs. The results can be found in Appendix A.1.

Finally, one may be concerned that the higher prices received by farmers with higher competition measures may arise from lower transport costs for farmers situated closer to rice mills rather than from the higher local competition among the mills. Column 4 addresses this by controlling for the distance between the farmer and the nearest mill. The results indicate that this concern is unfounded. First, the coefficient of interest is not affected by the inclusion of the additional control variable. Second, after controlling for the number of mills within 100km from the farmer, the distance between the farmer and the nearest mill does not have a significant effect on the farmer prices.

Overall, the results from the OLS regressions indicate a positive and significant relationship between local competition among rice mills and farmer prices. Farmers who are located in areas with a higher spatial density of rice mills tend to receive higher prices on average.

However, despite controlling for observable characteristics, there could still be concerns over other unobserved heterogeneity. Ex-ante, it is unclear how the bias would affect the coefficient of interest. For example, the direction of the bias from the quality of rice could be positive or negative. On the one hand, areas that are more suited for growing rice may produce higher quality rice and receive higher prices, leading to an upward bias. On the other hand, areas suitable for growing rice may choose to grow varieties of rice that give higher yield at the cost of lower quality and receive lower prices, leading to a downward bias. Given this, I turn to an instrumental variable strategy to establish a causal relationship between local competition among rice mills and farmer prices.

Taking advantage of the fact that farmers are small, I instrument for local competition among rice mills using the neighboring farmers' productivity. The intuition behind the instrument can be explained as follows. Mills' entry decisions depend on the profitability of the location. The total profit of any given mill depends on the quantity of rice that the mill sells, which is dictated by the quantity of rice that the mill can buy from the farmers. As a result, areas that produce a higher quantity of rice attract a larger number of rice mills, leading to a higher density of mills and higher competition. Since an area comprises a large number of farmers, the overall quantity of rice produced in an area depends on the aggregate productivity of farmers in that area. Therefore, the local competition among rice mills that farmer f experiences depends not only on farmer f's own productivity, but also on the productivity of the neighboring farmers. In other words, two farmers who have the same productivity may face different levels of local competition among the mills if the productivity of their neighbors is different.

Given this intuition, conditional on farmer f's own productivity, competition among rice mills can be instrumented using the productivity of the neighboring farmers. To further illustrate this, consider Figure 4. The center of the circle is the location of farmer f. Farmer f's own productivity is the productivity of the land within 50 km from the farmer, represented by the inner yellow circle. The neighboring farmers' productivity is the productivity of the land between 50 km and 100 km from the farmer, represented by the outer orange ring.[6] I employ the productivity in the outer ring as the instrument competition, controlling for the productivity in the inner circle. My instrument

[6] For alternative distance cutoff for the farmer's own productivity and the neighboring area's productivity, see Appendix A.2.

is similar to the instrument used by Macchiavello and Morjaria (2021). The first stage is given by:

$$COMP_{frct} = \alpha_0 + \underbrace{\alpha_1 A_f^{50-100}}_{Instrument} + \underbrace{\alpha_2 A_f^{0-50}}_{Control} + \alpha_3' Z_{ft} + \gamma_r + \gamma_c + \gamma_t + \mu_{ft}, \tag{3}$$

where A_f^{50-100} is the productivity of neighboring area that lies between 50 km and 100 km from the farmer, and A_f^{0-50} is the productivity of the land within 50 km from the farmer, which is used as a control for farmer f's own productivity. The predicted competition measure is then used in the second stage:

$$p_{frct}^f = \beta_0 + \beta_1 \widehat{COMP}_{frct} + \underbrace{\beta_2 A_f^{0-50}}_{Control} + \beta_3' Z_{ft} + \gamma_r + \gamma_c + \gamma_t + \varepsilon_{frct}. \tag{4}$$

The exclusion restriction is that, conditional on farmer f's own productivity and other controls included in the regression, the instrument only affects farmer prices through competition among the rice mills.[7] In the regressions, I proxy for the productivity using FAO crop-suitability-index and the area equipped for irrigation.[8]

Table 3 reports the 2SLS estimates. Columns 1 and 2 report the results using the competition measure constructed using equation (1). Column 1 shows that the instruments strongly correlate with competition, with an F-stat of 24.5. Column 2 reports results from the second stage. A one standard deviation increase in local competition measure increases the farmer prices by 7.7 %. The IV estimate (0.074) is about three times as large as the OLS estimate (0.023) in Column 2 of Table 2. The downward bias in the OLS estimate can be explained by the measurement error or the unobserved heterogeneity as outlined earlier. Columns 3 and 4 show that the results are robust to an alternative measure of competition. A one standard deviation increase in the number of mills

[7] This exclusion restriction would be violated if the productivity of the neighboring area correlates with other factors, such as distance to local roads, that correlate with farmer prices. Appendix A conducts additional robustness checks by controlling for terrain ruggedness and performing falsification exercises to show that the concern is unfounded.

[8] Specifically, I run the following first-stage:

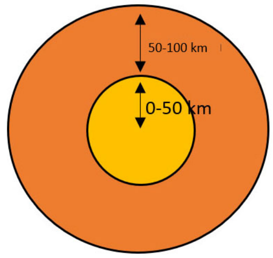
$$\begin{aligned} COMP_{frct} =& \alpha_0 + \underbrace{\alpha_1 Suit_f^{50-100} + \alpha_2 Suit_f^{50-100} \times Irri_f^{50-100} + \alpha_3 Irri_f^{50-100}}_{\text{Neighbors' productivity (instrument)}} \\ & + \underbrace{\alpha_4 Suit_f^{0-50} + \alpha_5 Suit_f^{0-50} \times Irri_f^{0-50} + \alpha_6 Irri_f^{0-50}}_{\text{Own productivity (controls)}} \\ & + \alpha_7' Z_{frct} + \gamma_r + \gamma_c + \gamma_t + \mu_{frct} \,, \end{aligned}$$

and the second stage:

$$\begin{aligned} p_{frct}^f = & \beta_0 + \beta_1 \widehat{COMP}_{frct} + \alpha_4 Suit_f^{0-50} + \alpha_5 Suit_f^{0-50} \times Irri_f^{0-50} + \alpha_6 Irri_f^{0-50} \\ & + \beta_3' Z_{ft} + \gamma_r + \gamma_c + \gamma_t + \varepsilon_{frct} \,. \end{aligned}$$

where $Suit_f^{50-100}$ and $Irri_f^{50-100}$ represent the crop-suitability-index and the area equipped for irrigation of the land between 50 km and 100 km from the farmer. Similarly, $Suit_f^{0-50}$ and $Irri_f^{0-50}$ represent the crop-suitability-index and the area equipped for irrigation within 50km from the farmer.

Figure 4: Instrument for Local Competition



Notes: This figure illustrates the instrument for local competition among rice mills. The orange outer ring represents the productivity of the neighboring farmers, specifically the productivity of the farmers situated between 50 to 100 km from the farmer. This is used as an instrument for local competition. The inner yellow circle represents the farmer's own productivity, which is used as a control variable.

within 100 km from the farmer leads to a 7.1 % increase in farmer prices. Further robustness checks can be found in Appendix A. I check for sensitivity of the instrument to different distance cutoffs, control for terrain ruggedness, and conduct falsification exercises. Overall, my results are robust to different specifications and alternative measures.

To summarize, this section provides empirical evidence of the spatial competition among rice mills. First, I show that there are variations in the spatial density of rice mills across Thailand and that the density of rice mills positively correlates with the quantity of rice harvested in the area. Second, I show that there is an active margin of entry and exit of rice mills, which impact the density of rice mills. Having done so, I provide reduced-form, causal evidence that local competition among rice mills, which is driven by the mills' spatial density, has significant effects on the farmer prices.

4 Model

I now develop a spatial model of trade that captures the empirical findings in the previous section. In the model, the market power of rice mills is captured through a Nash bargaining framework. The density of mills matters for farmer prices because it affects the farmer's threat point in the Nash bargaining process. Variations in the density of rice mills arise because mills make strategic entry and location decisions, which I model using a static entry game with incomplete information.

4.1 Setup

The economy consists of two types of agents, the farmers and the mills, who are distributed across the Euclidean space $\mathbb{R}^2$. There are F plots of land at different geographical locations. Each plot of land, f, is owned by a representative farmer. I index the farmer by their plot of land. There are M

Table 3: IV Results

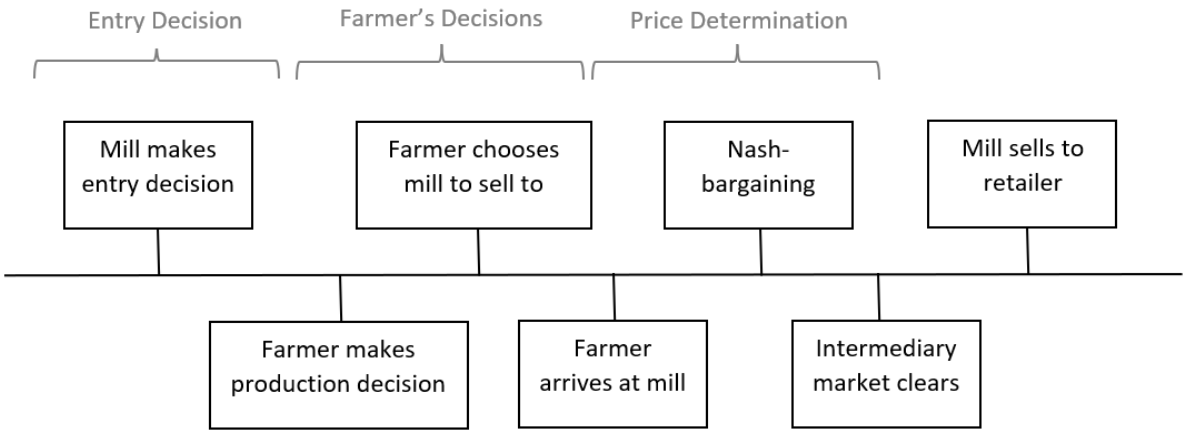
	COMP (std)	log(price)	No. of mills (std)	log(price)
COMP (std)		0.074***		
		(0.020)		
No. of mills (std)				0.069***
				(0.027)
$Suit^{50-100}$ (std)	0.308***		0.205***	
	(0.043)		(0.060)	
$Suit^{50-100} \times Irri^{50-100}$ (std)	0.183***		0.073**	
	(0.038)		(0.037)	
$Irri^{50-100}$ (std)	0.232***		0.413***	
	(0.062)		(0.150)	
$Suit^{0-50}$ (std)	0.070	0.013	0.113***	0.012
	(0.064)	(0.008)	(0.039)	(0.010)
$Suit^{0-50} \times Irri^{0-50}$ (std)	0.067	0.007	0.073**	0.008
	(0.051)	(0.009)	(0.035)	(0.010)
$Irri^{0-50}$ (std)	0.405***	-0.048***	0.254***	-0.037**
	(0.124)	(0.015)	(0.064)	(0.015)
First-stage F-stats	24.514		38.837	
Hansen's p-value		0.870		0.222

Notes: This table reports the IV estimates of equations (3) and (4). COMP (std) is the standardized competition measure constructed using equation (1). No. of mills (std) is the standardized number of mills within 100 km from the farmer. All regressions include controls for quantity of crop sold, province-level output, distance to Bangkok, and fixed effects for year, month-of-highest-sales, region, and crop-type. All regressions use data from 2008 to 2015. Standard errors are adjusted to allow for spatial clustering as in Conley (1999), with a bandwidth of 1.5 degree using Bartlett kernel. *$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

potential entrants of rice mills. Let $\mathcal{M}$ be a set of mills that are operating. Each mill m is identified by the co-ordinates of their geographical location. Geography matters since the transportation of rice across space is subject to iceberg trade costs. $\tau_{ij} \geq 1$ units of rice must be shipped in order for one unit of rice from i to arrive at j. I assume that trade costs are symmetric (i.e. $\tau_{ij} = \tau_{ji}$) and that triangular inequality holds (i.e. $\tau_{ij} \leq \tau_{ik} \times \tau_{kj}$).

The timing of events is outlined in Figure 5. First, the mills make their entry and location decisions. The farmers then choose the quantity of rice to grow. After the rice has been harvested, the farmers choose which mill to sell their rice to. Once the farmers have arrived at the mill, the farmer prices are determined through Nash-bargaining. After the intermediary markets clear, the mills then sell the rice to the retailers.

Figure 5: Timing of Events



4.1.1 Farmers

Farmer produces rice according to the production function:

$$y_f = A_f \overline{H}_f^{\gamma} \overline{L}_f^{\alpha} X_f^{\beta} \ , \tag{5}$$

where A_f is the total factor productivity of the land owned by farmer f, H_f is the amount of land, L_f is the quantity of labor, and X_f is the quantity of intermediate inputs. I assume that the production function has constant returns to scale, so $\alpha + \beta + \gamma = 1$. The amount of land and labor available to the farmer is fixed. I assume that the farmers are price takers in the intermediate input and the labor markets. Farmers optimally choose the quantity of input to maximize their income, producing the profit maximizing output y_f^*.

After the production has taken place, the farmer chooses mill m to sell their crop to in order to maximize their income. Since transporting rice from location f to m is subject to iceberg trade cost τ_{fm}, their maximization problem is given by:

$$\max_{m \in \{\mathcal{M}\}} \frac{p_m^f y_f^*}{\tau_{fm}} \ , \tag{6}$$

where p_m^f is the farmer price at mill m, and y_f^* is the profit-maximizing output.

4.1.2 Mills

Mills are price takers in the retail sector. Mill m buys a unit of rice from the farmer at a price p_m^f and sells it to the retailer at an exogenously given retail price p_m^r. Mills can have different retail prices, depending on their locations. I assume that there is no marginal cost; therefore, the per-unit profit for the mill from trading rice is $p_m^r - p_m^f$.

4.1.3 Price Determination

Price in the intermediary markets is determined by Nash bargaining as in Chatterjee (2020). Once farmer f has arrived at mill m, the initial transport cost from f to the m is sunk. The farmer could either sell their rice to mill m and receive p_m^f per unit of rice, or travel to another mill k and receive $\frac{p_k^f}{\tau_{mk}}$ since it is costly to transport crop from m to k. Therefore, for a farmer who has already arrived at mill m, the outside option is:

$$\underline{p}_m = \max_{k \in \mathcal{M}-\{m\}} \left\{ \frac{p_k^f}{\tau_{mk}} \right\} , \tag{7}$$

which is the highest value that farmers could get if they travel to another mill and sell their crop there. Farmer's outside option forms their threat point in the Nash bargaining problem. The outside option of the mill is 0.

The farmer price is determined by the Nash bargaining solution, which solves the following maximization problem:

$$\max_{p_m^f} \quad (p_m^f - \underline{p}_m)^{\delta} (p_m^r - p_m^f)^{1-\delta} , \tag{8}$$

where δ is the bargaining power of the farmer. The left-hand side bracket is the difference between the farmer price and the farmer's threat point. The right-hand side bracket is the per-unit profit that rice mills receive from this transaction. Solving the Nash bargaining problem provides an expression for the farmer price at each mill in terms of the threat point and the retail price:

$$p_m^f = (1-\delta)\underline{p}_m + \delta p_m^r . \tag{9}$$

We have one such equation for all $m \in \mathcal{M}$. The threat point $\underline{p}_m$ is a function of the farmer prices at all other locations, as defined by (7). Therefore, the equilibrium is a Nash-in-Nash solution, consisting of a vector of farmer prices that solves the fixed-point mapping (9). The fixed-point problem (9) is a contraction mapping and therefore has a unique solution.[9] Note that all farmers arriving at mill m have the same outside option $\underline{p}_m$. Therefore, all the farmers arriving at the same mill receive the same price. In other words, there is only one farmer price at each mill.

4.1.4 Entry Decision

Following the framework in Seim (2006), I model the entry and location decisions of rice mills as a static game of incomplete information. There are M potential entrants who simultaneously choose whether to enter and where to locate from a set of L possible locations. Location l is defined as a non-overlapping grid cell containing all points $\Big(x \in [\underline{x}_l, \overline{x}_l], y \in [\underline{y}_l, \overline{y}_l]\Big)$. Each mill can only choose to operate in one location in any given period. Upon entering location l, mill m randomly draws

[9] See Appendix B for details.

$x_m \sim U(\underline{x}_l, \overline{x}_l)$, $y_m \sim U(\underline{y}_l, \overline{y}_l)$. Intuitively, if one think of L as a set of possible districts; then, a mill can choose which district they want to locate in, but the exact co-ordinates within a district that they get depends on many exogenous factors such as the availability of land; therefore, each mill has their own unique co-ordinates. All distance calculations in the model use the mills' actual co-ordinates.

The variable profit of mill m in location l is $(p_l^r - p_{ml}^f)Q_{ml}$, where the retail price p_l^r is exogenously given, the farmer price p_{ml}^f is determined from the Nash bargaining problem, and Q_{ml} is the quantity of rice that mill m buys from farmers. I assume that all the mills in the same location receive the same retail price. Let the spatial distribution of mills be represented by a vector $N = \{n_1, n_2, ..., n_L\}$, where n_l is the number of mills in location l. Conditional on N, the expected variable profit of a mill located in location l is given by:

$$vp_l^e(N, P^r, A, \overline{H}, \overline{L}) = \mathbb{E}\left[(p_l^r - p_{ml}^f)Q_{ml}\right] \ , \tag{10}$$

where $P^r = \{p_1^r, p_2^r, ..., p_L^r\}$ is a vector of exogenously given retail price across locations and $A = \{A_1, A_2, ..., A_F\}$ is a vector representing the farmers' productivity, $\overline{H} = \{\overline{H}_1, \overline{H}_2, ..., \overline{H}_F\}$ is a vector representing the farmers' land endowment, and $\overline{L} = \{\overline{L}_1, \overline{L}_2, ..., \overline{L}_F\}$ is a vector representing the farmers' labor endowment. Note that $A, \overline{H}$, and $\overline{L}$ are all exogenous characteristics of the farmers that determine their output; therefore, for ease of notation, I omit $\overline{H}$ and $\overline{L}$ from the subsequent writing of the function vp_l^e.

The total profit is the variable profit minus the entry cost. Mills make location decisions based on total profit in each location. Conditional on N, the expected total profit of mill m in location l is:

$$\pi_{ml}^e(N, P^r, A) = vp_l^e(N, P^r, A) - ec_{ml} \ . \tag{11}$$

The cost of entry into location l for mill m is given by:

$$ec_{ml} = EC_l - \varepsilon_{ml}^{EC} \ , \tag{12}$$

where EC_l is the location-specific entry cost that is common for all mills and ε_{ml}^{EC} is the mill and location-specific cost that is private information of mill m. Because mills possess private information about their cost of entry, N is unknown ex-ante. Each mill has to form beliefs about their rivals' choices of locations. Since the mills are homogeneous apart from their private information, all mills have the same belief about all their rivals. Let $\overline{\Psi} = \{\overline{\psi}_1, \overline{\psi}_2, ..., \overline{\psi}_L\}$ be the mill's belief about their rivals' location strategy, where $\overline{\psi}_l$ is the probability a rival will choose to enter location l. The expected total profit for a mill located in location l without conditioning on N is:

$$\mathbb{E}_N\left[\pi_{ml}^e(N, P^r, A)\right] = \widetilde{vp}_l^e(\overline{\Psi}, P^r, A) - EC_l + \varepsilon_{ml} \ , \tag{13}$$

where $\widetilde{vp}^e_l(\overline{\Psi}, P^r, A) = \mathbb{E}_N\left[vp^e_l(N, P^r, A)\right]$ is the expected variable profit over the distribution of N. Note that since mills never know the realized variable profit vp_l prior to entering, they make all entry decisions based on $vp^e_l(N, P^r, A)$. Therefore, to avoid confusion, I henceforth refer to $vp^e_l(N, P^r, A)$ as the variable profit and $\widetilde{vp}^e_l(\overline{\Psi}, P^r, A)$ as the expected variable profit.

Given its beliefs, $\overline{\Psi}$, a mill chooses the entry-location decision that maximizes its expected profit. The best response of mill m is to locate in location l if:

$$\widetilde{vp}^e_l(\overline{\Psi}, P^r, A) - EC_l + \varepsilon_{ml} \geq \widetilde{vp}^e_{l'}(\overline{\Psi}, P^r, A) - EC_{l'} + \varepsilon_{ml'} \quad \forall l' \neq l\ . \tag{14}$$

Following the IO literature on entry of firms (Seim, 2006; Zhu and Singh, 2009; Datta and Sudhir, 2013; Aguirregabiria and Vicentini, 2016), I assume that the private information shocks are independently and identically distributed across mills and locations with type 1 extreme value distribution.[10] The probability that the best response of a mill is to locate in location l can then be expressed as multinomial logit probabilities:

$$\psi_l(\overline{\Psi}) = \frac{exp\left\{\lambda\left[\widetilde{vp}^e_l(\overline{\Psi}, P^r, A) - EC_l\right]\right\}}{1 + \sum\limits_{l'=1}^{L} exp\left\{\lambda\left[\widetilde{vp}^e_{l'}(\overline{\Psi}, P^r, A) - EC_{l'}\right]\right\}}\ , \tag{15}$$

where 1 in the denominator accounts for the option for a mill to not enter the market. λ is the scale parameter the captures the dispersion of the private information shock. The lower the value of λ, the higher the dispersion and the more disconnected the entry decision is to the expected profit.

The equilibrium is a symmetric Bayesian Nash equilibrium in which the mills' beliefs are consistent with other mills' best responses. Given that mills are homogenous apart from their private information shocks, in equilibrium, all mills have the same strategy. This gives us a system of L equations with L unknown that defines a fixed point mapping in the space of vector of entry-location probabilities:

$$\psi^*_l(\Psi^*) = \frac{exp\left\{\lambda\left[\widetilde{vp}^e_l(\Psi^*, P^r, A) - EC_l\right]\right\}}{1 + \sum\limits_{l'=1}^{L} exp\left\{\lambda\left[\widetilde{vp}^e_{l'}(\Psi^*, P^r, A) - EC_{l'}\right]\right\}}\ . \tag{16}$$

By Brouwer's fixed point theorem, an equilibrium exists. However, the equilibrium is not necessarily unique. Seim (2006) shows that there is a unique solution if the locations are not too homogeneous and if the degree of competition decreases with distance. This means that in my model, the iceberg trade cost must be sufficiently large. I check ex-post through simulations that the equilibrium appears to be unique.

[10] This assumption is computationally attractive because it provides a closed-form expression for the mill's best response function. However, it comes at a cost since it implies that there is no spatial correlation in the private information.

4.2 Equilibrium Definition

Given the parameters $\alpha, \beta, \gamma, \delta$, and λ, iceberg trade cost function $\{\tau\}$, endowments $\{A_f, \overline{H}_f, \overline{L}_f\}_{f\in\{1,...,F\}}$, intermediate input price w^X, and a vector of retail prices P^r, location-specific component of entry cost $\{EC_l\}_{l\in\{1,...,L\}}$, and number of potential entrants M, an equilibrium is a set of entry location probabilities $\Psi^* = \{\psi_1^*, ..., \psi_L^*\}$, farmer prices $P^f = \{p_1^f, ..., p_{|\mathcal{M}|}^f\}$, intermediate input choices $\{X_1, ..., X_F\}$, and the mill choice for each farmer $\{\mu_1, ..., \mu_F\}$ such that:

1. Farmers optimally choose the quantity of intermediate input to maximize their incomes:

$$\max_{X_f} \quad \frac{p_{\mu_f}^f}{\tau_{f\mu_f}} A_f \overline{H}_f^\gamma \overline{L}_f^\alpha X_f^\beta - w^X X_f \qquad \forall f \tag{17}$$

2. Farmers make an optimal choice on which mill to sell their output:

$$\mu_f = \arg\max_{m\in\mathcal{M}} \frac{p_m^f y_f^*}{\tau_{fm}} \qquad \forall f \tag{18}$$

3. Farmer prices across all mills are the Nash-in-Nash equilibrium solution:

$$p_m^f = (1-\delta) \max_{k\in\mathcal{M}-\{m\}} \left\{ \frac{p_k^f}{\tau_{mk}} \right\} + \delta p_m^r \qquad \forall m \in \mathcal{M} \tag{19}$$

   which combines equations (9) and (7) into one equation.

4. Mill's entry-location probabilities, Ψ, across all locations satisfy the symmetric Bayesian Nash equilibrium given by equation (16).

4.3 Equilibrium Computation

I solve the model backward from the timeline in Figure 5. First, conditional on the retail prices and the number of mills at each location, I solve the Nash bargaining problem for the farmer prices as given by equation (19). Next, given the farmer price at each mill, I solve each farmer's optimal choice of mill as given by equation (18). Subsequently, I know the price that each farmer receives and can solve for each farmer's optimal choice of intermediate input, which gives me the quantity of rice each farmer produces. Finally, given the retail prices, farmer prices, farmers' output, and farmers' optimal mill choice, I solve for the mills' expected profit then solve for the Bayesian Nash equilibrium, which determines mills' entry-location decisions. In what follows, I describe how I compute the Bayesian Nash Equilibrium.

4.3.1 Bayesian Nash Equilibrium Computation

The Bayesian Nash equilibrium from the mills' entry problem is a solution to the fixed point problem in equation (16). Solving this requires knowledge of the expected variable profit, $\widetilde{vp}_l^e(\overline{\Psi}, P^r, A)$. Let $\mathbb{S} = \{N_1, ..., N_{|\mathbb{S}|}\}$ be a set of all possible combinations of the number of mills in each location. In theory, one can calculate the expected variable profit, $\widetilde{vp}_l^e(\overline{\Psi}, P^r, A)$, by computing the variable profit, $vp_l^e(N, P^r, A)$, for each $N_s \in \mathbb{S}$ and calculate the expected variable profit conditional on $\overline{\Psi}$, as $\widetilde{vp}_l^e(\overline{\Psi}, P^r, A) = \mathbb{E}_N [vp_l^e(N, P^r, A)]$. To do so, one would have to keep track of $|\mathbb{S}| \times L$ variable profits corresponding to each N to compute the expected variable profit. However, it is computationally infeasible to enumerate all possible configurations of N and solve for the corresponding $vp_l^e(N, P^r, A)$. To illustrate the extent of this dimensionality challenge, consider a small entry game with only 10 locations and 100 potential entrants. Even in such a small entry game, we already have $|\mathbb{S}| > 10^{13}$; this means that, in order to compute the exact expected variable profit, one would need to keep track of over $10^{13} \times L$ parameters.[11] Even in such a small entry game, the dimensionality of the space of mills' actions is already too large to be computationally feasible. In the data, I observe an average of 1,011 mills per year. Therefore, the number of potential entrants that I need to consider must be over a thousand, posing a much larger dimensionality challenge than in the illustrative example above.

The literature generally approaches this computational issue by approximating the equilibrium. For example, in order to deal with the dimensionality problem in a dynamic entry game, Aguirregabiria and Vicentini (2016) approximate the best response functions of firms using an interpolation function that is second order polynomial in the number of firms and distances between firms. In the same spirit, I compute the approximated Bayesian Nash Equilibrium by approximating the variable profit function. As explained earlier, for a given set of retail prices and productive capacity of the farmer, the only information I need to compute the variable profit is the number of mills in each location. Since P^r and A are exogenously given, I write the expected variable profit conditional on N, $vp_l^e(N, P^r, A)$, from (10) as a function of N i.e. $vp_l^e(N, P^r, A) = F_{P^r,A,l}(N)$. For ease of notation, I omit P^r and A from subsequent writing of the variable profit and approximation functions; however, please note that the variable profit and the approximation function are specific to given P^r and A. I approximate the function $F_l(N)$ as a second order polynomial in the number of mills

[11] $|\mathbb{S}|$ is computed as follows. There are 10 locations, and potential entrants can choose not to enter, resulting in 11 options for potential entrants to choose from. Since mills are homogeneous apart from the private information shocks, the ordering does not matter. Therefore, this is equivalent to an unordered sampling with replacement; we are choosing from a set of 11 a hundred times such that repetition is allowed and such that order does not matter. The number of possible spatial configurations of N in this small game is $\binom{110!}{10!100!} > 10^{13}$. More generally, for a game with L locations and M potential entrants, the number of possible realizations of N is $\binom{(L+M)!}{L!M!}$.

in each location:[12]

$$vp_l^e(N) \approx \beta_l^{(0)} + \sum_{k=1}^{L} \beta_{kl}^{(1)} n_k + \sum_{k=1}^{L} \beta_{kl}^{(2)} n_k^2 + \nu_l \quad \forall l. \tag{20}$$

By doing so, I am able to reduce the dimension of the problem from $|\mathbb{S}| \times L$ down to $(2L+1)L$.[13]

I now describe in detail the procedure I use to obtain the approximation function for the variable profit. Let $S = \{N_1, N_2, ..., N_{|S|}\}$ be a subset of $\mathbb{S}$, which is a set of all possible N. Given P^r and A, I compute the simulated variable profit $vp_l^e(N_s, P^r, A)$ for each $N_s \in S$. This means that I have a simulated dataset consisting of $|S|$ observations. For each $l \in L$, I obtain the approximation function (20) by running the following OLS regression on the simulated dataset:

$$vp_{ls}^e = \beta_l^{(0)} + \sum_{k=1}^{L} \beta_{kl}^{(1)} n_{ks} + \sum_{k=1}^{L} \beta_{kl}^{(2)} n_{ks}^2 + \nu_{ls} \ . \tag{21}$$

Having obtained $\beta_l^{(0)}, \{\beta_{1l}^{(1)}, \beta_{2l}^{(1)}, ..., \beta_{Ll}^{(1)}\}$, and $\{\beta_{1l}^{(2)}, \beta_{2l}^{(2)}, ..., \beta_{Ll}^{(2)}\}$, for a given belief $\overline{\Psi}$, I can approximate for the expected variable profit as:

$$\widetilde{vp}_l^e(\overline{\Psi}) \approx \beta_l^{(0)} + \sum_{k=1}^{L} \beta_{kl}^{(1)} \mathbb{E}[n_k] + \sum_{k=1}^{L} \beta_{kl}^{(2)} \mathbb{E}[n_k^2] \ , \tag{22}$$

where

$$\mathbb{E}[n_k] = \overline{\psi}_k \cdot (M-1) + \mathbb{I}_{k=l} \ , \tag{23}$$

$$\mathbb{E}[n_k^2] = Var(n_k) + \mathbb{E}[n_k]^2 \ . \tag{24}$$

For a mill considering whether to enter location l, the expected number of mills at location k when $k \neq l$ is the probability that a rival would choose to enter location k, $(\overline{\psi}_k)$, multiplied by the number of potential rivals $(M-1)$. If $k = l$, then the mill includes itself into the expected number of mills, resulting in $\mathbb{E}[n_l] = \overline{\psi}_l \cdot (M-1) + 1$. The approximated Bayesian Nash equilibrium is therefore the solution to the fixed point problem:

$$\psi_l^*(\Psi^*) = \frac{exp\left\{\lambda\left(\beta_l^{(0)} + \sum_{k=1}^{L} \beta_{kl}^{1} \mathbb{E}[n_k] + \sum_{k=1}^{L} \beta_{kl}^{(2)} \mathbb{E}[n_k^2] - EC_l\right)\right\}}{1 + \sum_{l'=1}^{L} exp\left\{\lambda\left(\beta_{l'}^{(0)} + \sum_{k=1}^{L} \beta_{kl'}^{1} \mathbb{E}[n_k] + \sum_{k=1}^{L} \beta_{kl'}^{(2)} \mathbb{E}[n_k^2] - EC_{l'}\right)\right\}} \ . \tag{25}$$

[12] This approach is equivalent to assuming that agents are boundedly rational in their perception of the variable profit as has been done in the literature (Krusell and Smith, 1998; Dingel and Tintelnot, 2021).

[13] By approximating the variable profit as a second order polynomial in the number of mills in each location, I am making the following assumptions: 1) the marginal effect of n_k on vp_l^e is $\beta_{kl}^{(1)} + \beta_{kl}^{(2)} n_k$, and 2) the marginal effect of n_k is not affected by n_j, $j \neq k$.

A critical question is how well the approximated equilibrium matches up to the true equilibrium. To validate the approximation method, I apply the method to the same entry problems but with a much smaller number of locations and a much smaller number of potential entrants. I consider the case with four locations and 50 potential entrants. At this scale, the problem is small enough to compute the Bayesian Nash equilibrium using the exact expected variable profit, which is computed over all spatial configurations of N. Table 4 shows the resulting expected number of mills at each location. We can see that the approximated equilibrium is very similar to the exact equilibrium. I repeat the exercise with five locations. The approximation method yields a very similar equilibrium to the exact equilibrium across various specifications. The key takeaway is that the approximation method provides a reasonably good approximation to the exact solution to the Bayesian Nash equilibrium.

Table 4: Exact vs Approximated Bayesian Nash Equilibrium

	L = 4, M = 50		L = 5, M = 40		L = 5, M = 50	
	Exact	Approx.	Exact	Approx.	Exact	Approx.
E(N)	9.39	9.66	5.62	5.76	7.15	7.29
	11.58	11.44	8.27	8.21	9.91	9.96
	11.74	11.46	7.17	7.04	8.82	8.74
	7.62	7.72	7.89	7.90	9.57	9.57
			4.88	4.94	6.41	6.41

Notes: This table reports the equilibrium expected number of mills in three different entry games. L denotes the number of locations and M denotes the number of potential entrants. Each row shows the expected number of mills in each location. "Exact" column reports the true equilibrium. "Approx" column reports the approximated equilibrium, where the expected variable profit is approximated using equation (20).

5 Estimation

In this section, I estimate key parameters in the model. To take the model to the data, I partition Thailand into 0.5 degrees by 0.5 degrees grid, creating 180 locations where mills can choose to enter (L = 180). To represent the farmers, I divide Thailand into 0.1 degrees by 0.1 degrees grid, resulting in 1,175 farmer locations (F = 1,175). I calibrate the farmers' production function using data from the agricultural survey. I then estimate key parameters in the model in two steps. First, taking N in the data as given, I jointly estimate the bargaining power δ and the trade cost in the Nash bargaining problem using method of simulated moments. Second, using the estimated Nash-bargaining parameters and the calibrated production function, I estimate parameters in the entry problem using the nested fixed point maximum likelihood algorithm.

5.1 Nash-bargaining parameters: δ and τ

Equilibrium in the Nash-bargaining problem is defined by the Nash-in-Nash equilibrium equation (19). I assume that trade cost between m and k takes the functional form:

$$\tau_{mk} = 1 + \phi d_{mk} \ , \tag{26}$$

where d_{mk} is the geodesic distance between m and k calculated using m and k actual co-ordinates.[14] By definition, $\tau_{mk} = 1$ if $m = k$. Nash-in-Nash equilibrium equation (19) then becomes:

$$p^f_m = (1-\delta) \max_{k \in \mathcal{M}-\{m\}} \left\{ \frac{p^f_k}{1+\phi d_{mk}} \right\} + \delta p^r_m \qquad \forall m \in \mathcal{M} \ , \tag{27}$$

The two parameters I need to estimate for the Nash-bargaining problem are the trade cost parameter ϕ and the bargaining power parameter δ.

Given that the farmer prices across different mills are interrelated through the outside options, I estimate the Nash-bargaining parameters $\Theta^{NB} = (\delta, \phi)$ by the method of simulated moments. I use the annual average of the prices of white rice and sticky rice between 2008-2018. I search over Θ^{NB} that minimizes the distance between the simulated moments and the data moments, achieving:

$$\hat{\Theta}^{NB} = \underset{\Theta^{NB}}{\arg\min} \left(\varphi^d - \varphi^s(\Theta^{NB}) \right) W \left(\varphi^d - \varphi^s(\Theta^{NB}) \right)' \ , \tag{28}$$

where φ^d represents the data moments, $\varphi^s(\Theta^{NB})$ represents the simulated moments, and W is the inverse of the variance-covariance matrix.[15] For each iteration of simulation at a new value of Θ^{NB}, I solve for the Nash-in-Nash equilibrium for each year and rice type between 2008-2018, taking the number of mills in the data as given.

The moments I choose to identify δ and ϕ come from an auxiliary regression and the distribution of farmer prices. First, following Chatterjee (2020), I adopt an auxiliary linear regression model that closely reflects the equilibrium equation (27):

$$p^f_{mt} = \beta_0 + \beta_1 \left(\max_{k \text{ s.t. } d_{mk} < 100km} p^f_{kt} \right) + \beta_2 d_{mk} + \beta_3 p^r_{mt} + \varepsilon_{mt} \tag{29}$$

[14] This functional form is similar to Chatterjee (2020). However, whereas Chatterjee (2020) explicitly includes and exogenous trade cost term in the functional form, in this paper, the randomness in trade cost is driven by the realization of the mill's co-ordinates.

[15] I calculate W by bootstrapping the initial dataset. Specifically, I resample with replacement from the initial dataset 1,000 times and calculate the corresponding moments from each sample. I then calculate the covariance between the bootstrapped moments.

where m indexes the mill, t indexes the year, and d_{mk} is the distance between mill m and mill k.[16] The moments I target are the coefficients β_1, β_2, and β_3. Second, I also match the distribution of farmer prices by targeting the mean, the 5th, the 25th, the 50th, the 95th percentile of the farmer prices. Although all parameters are jointly estimated and there is no one-to-one mapping between the parameters and the moments, some moments are more sensitive to certain parameters than others. Intuitively, δ is identified from β_1 and β_3 which capture the correlation between farmer price at mill m and farmer price at a competing mill and the correlation between farmer prices and retail prices respectively, whereas ϕ is identified through β_1 and β_2, which captures how distances between mills correlate with the farmer prices.

Note that to estimate the auxiliary regression (29), I make the following assumptions about the data. First, since I only observe the prices that farmers received and not which mills they sold the rice to in my dataset, in order to run this auxiliary linear regression, I assume that if a mill is situated in the same grid cell as the farmer, then the farmer sold rice to that mill and the reported price that the farmer received is p^f_{mt}, the price that the mill gave the farmer. Second, I only observe retail prices at the province level. Therefore, I assume that all mills located in the same province receive the same retail price. For province-year with missing retail prices, I linearly interpolate the retail prices from adjacent provinces.

Table 5 reports the estimated values of the parameters. The bargaining power of farmer δ is estimated to be 0.39, and the trade cost parameter ϕ is estimated to be 0.84. The estimated trade cost parameter is relatively high since it also captures elements of trade costs beyond the distance-related transport cost. For instance, in transporting rice to sell to another mill, the farmer faces the risk that they may not make it in time before the other mill closes their buying window, meaning the farmer may have to make another trip on the next day. Additionally, it is difficult for farmers to arrange transportation to another mill in practice since the agreement to transport rice from the farm to the rice mill is often done on a leg basis. Table 6 displays the moments from the data and simulated data. Column 1 provides the moments calculated from the data, which are the moments that I target, and column 2 displays the moments generated by the model.

I check for the goodness of fit in the following ways. Figure 6(a), shows the binned scatter plot of the log of the farmer prices that I observe in the data and the log of farmer prices generated in the model. The farmer prices simulated in the model strongly correlate with the data. The coefficient from a linear fit without an intercept term is 0.984. Figure 6(b) displays the histogram of the log of farmer prices in the data and the farmer prices simulated in the model. Overall, the model is able to capture the distribution of farmer prices, although the model generates a slightly

[16] I do not use the reduced form IV regression (4) as my auxiliary regression model because it does not help with identification of δ and ϕ since the reduced form regression (4) captures all the competition effect through the $COMP$ measure. Instead, I will use the coefficient of the $COMP$ measure as an untargeted moment to assess the goodness of fit.

Table 5: Estimated Parameters from MSM

Parameter	Point Estimate	SE
δ	0.3876	0.0010
ϕ	0.8352	0.0058

Notes: This table reports the point estimates of the δ, the bargaining power of the farmer, and ϕ, the trade cost parameter using the method of simulated moments. Parameters are estimated using farmer prices between 2008-2018.

Table 6: MSM Targeted Moments

	Data	Model
β_1	0.840	0.889
β_2	-0.014	-0.004
β_3	0.013	0.049
Mean	10.829	10.948
25th	8.600	9.452
50th	10.750	10.825
75th	12.828	12.489

Notes: This table reports the targeted moments used in MSM to estimate parameters in the Nash bargaining problem. β_1, β_2, and β_3 are coefficients from the auxiliary regression. (29). The last four rows show the mean, the 25th, the 50th, and the 75th percentile of farmer prices.

wider variation in price compared to the data.[17] I further check how well my model is able to match the empirical results I presented in Section 3. Since I did not use the IV specification (4) in my estimation, I use β_1, the coefficient that captures how local competition impacts farmer prices, from the IV regression as my untargeted moment. An equivalent IV regression using the simulated data yields a coefficient of 0.079, which is relatively close to the coefficient of 0.074 in the data.

5.2 Parameters in Entry Problem: λ and EC

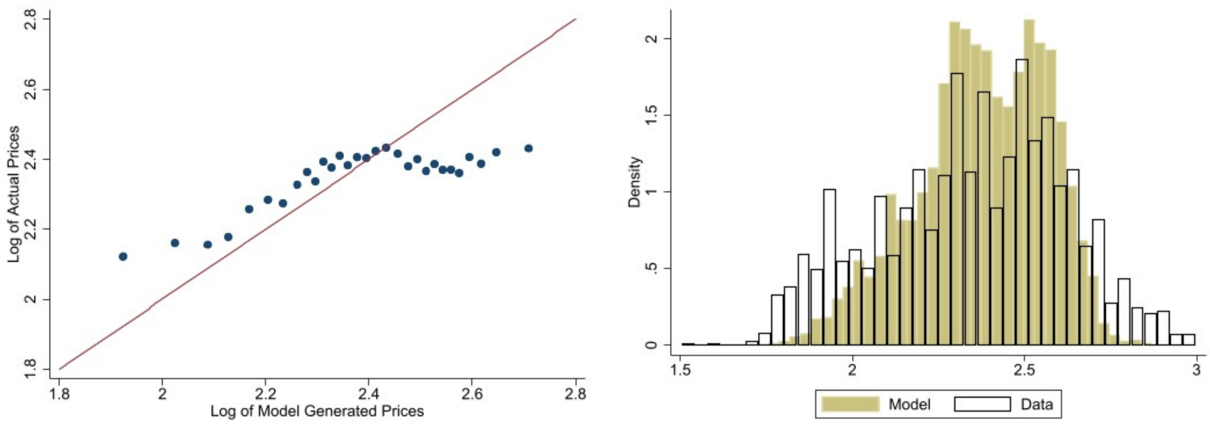
The equilibrium in mills' entry-location problem is given by the Bayesian Nash equilibrium equation (16). The parameters I need to estimate are the scale parameter, λ, and the common knowledge component of entry costs, EC_l. I assume that EC_l takes the following specification:

$$EC_l = \theta_0^{EC} + \theta_1^{EC} ruggedness_l + \theta_2^{EC}\ \%\ of\ in\ season\ rice_l + \theta_3^{EC} population\ density_l + \gamma_r \quad (30)$$

[17]Farmer prices in the data can be more compressed than the model-generated prices because of several reasons. First, farmers could have other outside options not captured by the model, such as keeping the rice at home. Likewise, mills may also have an upper bound on the prices they are willing to give to farmers, beyond which they consider to be unprofitable to buy rice from the farmers. In addition, some rice mills may choose to buy only certain rice varieties, which affects the degree of local competition in a way that is not captured in the model.

Figure 6: MSM Goodness of Fit

(a) Binned Scatter Plot of Farmer Prices (b) Distribution of Farmer Prices



Notes: Panel (a) shows a binned scatter plot of the log of farmer prices observed in the data against the log of farmer prices simulated from the model using estimates from the MSM. The red line is the 45-degree line. Panel (b) shows a histogram of the log of farmer prices in the data and the log of the model generated farmer prices. Both panels use farmer prices of white and sticky rice between 2008-2018.

Table 7: MSM Untargeted Moment

	log(price)	
	Data	Model
COMP (std)	0.074	0.079

Notes: This table reports the coefficient for the standardized competition measure from the IV regression (4). The first column reports the estimated coefficient from the data and the second column reports the estimated coefficient from the model generated data. Note that because data from the model is generated at an annual frequency, I exclude the fixed effects for the month-of-highest-sales from the regression using the model simulated data.

where $ruggedness_l$ is the ruggedness of the location, $\%\ of\ in\ season\ rice_l$ is the percentage of rice in the area that is grown in season, $population\ density_l$ is the population density of the location, and γ_r is the region fixed-effect.[18] Given the specification of EC_l, the set of parameters I need to estimate for the entry problem $\Theta^{EC} = \{\lambda, \theta_0^{EC}, ..., \theta_3^{EC}, \gamma_{r\in R}\}$.

Estimation of Θ^{EC} proceeds via nested fixed-point algorithm. Given $\psi_l^*\big(\Theta^{EC}, X_t\big)$, the probability

[18] I choose the variables in the specification for EC_l the following reasons. Terrain ruggedness matters for the cost of doing busing, as Nunn and Puga (2012) have stated, *"geographical ruggedness is an economic handicap,... making it more expensive to do business."* The percentage of rice grown in season likely affects the fixed operating cost of the mill since it indicates whether farmers in the area grow rice consistently throughout the year or only during certain time period. If farmers only grow rice during a specific part of the year, mills in the area will likely have to shut down during some part of the year. The percentage of rice grown in season captures the costs arising from mills' inability to conduct businesses continuously. The population density captures variations in the cost of land. The region-fixed effects captures variations in the entry cost across the regions.

that the best response of a mill is to enter location l as given by the equilibrium probability (16), the number of mills at a location follows a multinomial distribution. The probability that we would observe $\{n_{1t}, ..., n_{Lt}\}$ number of mills at time t is:

$$Pr\Big(n_{1t}, ..., n_{Lt} | \Psi^*\big(\Theta^{EC}, X_t\big)\Big) = \frac{M!}{n_{1t}! \cdots n_{L+1,t}!} \prod_{l=1}^{L+1} \psi_l^*\big(\Theta^{EC}, X_t\big)^{n_{lt}} , \tag{31}$$

where X_t are the variables needed to determine the expected variable profit. Assuming there is no unobserved location heterogeneity, the log-likelihood function is then given by:

$$l\left(\Psi^*\big(\Theta^{EC}, X_t\big)\right) = \sum_{t=1}^{T} \sum_{l=1}^{L+1} \left\{ n_{lt} \ln \psi_l^*\big(\Theta^{EC}, X_t\big) \right\} . \tag{32}$$

The log-likelihood function is summed up to $L+1$ to account for the mills' option to not enter altogether, where $\psi_{L+1}^*\big(\Theta^{EC}, X_t\big) = 1 - \sum_{l=1}^{L} \psi_l^*\big(\Theta^{EC}, X_t\big)$ and $n_{L+1t} = M - \sum_{l=1}^{L} n_{lt}$. Calculation of log-likelihood function therefore requires knowledge of M, the number of potential entrants, which is unobserved. Following the literature, I fix the number of potential entrants, M, to be an exogenous value, setting $M = 2,000$.[19]

The nested fixed point maximum-likelihood estimation proceeds as follows. For a given set of variables X_t which determines the variable profits, I randomly sample a subset S of size $60,000$ from all possible configurations of N and compute the simulated variable profit for each $N_s \in S$.[20] I then calculate the approximation function for the variable profit according to (20). Having done so, or a given set of parameter values Θ^{EC} and approximated variable profit function, I solve for an approximated fixed point solution to the Bayesian Nash equilibrium (25). This Bayesian Nash equilibrium is then nested into the maximum-likelihood procedure to estimate the parameters Θ^{EC}.

To calculate the variable profit, I need data on the quantity of crop output. Data on the quantity of rice harvested at the district level is only available between 2012-2018. Therefore, I restrict my sample size to that time period. Additionally, I use the retail price of white rice in my estimation since white rice is more widely grown across Thailand.

Table 8 reports the estimated parameters. The goodness of fit can be seen in Figure 7. Figure 7(a) displays the binned scatter plot of the number of mills in each location that I observe in the

[19]Since the number of potential entrants is unobserved, it is standard in the literature to set this to an exogenous number. For instance, one of the specifications that Seim (2006) use is to assume that there is an entrant pool of 50 firms and the other specifications is to set the number of potential entrants such that 50% of the potential entrants enter the market. since the number of mills per year I observe on average is 1,011 and the highest number of mills observed in the data is $1,084$, I set M to be $2,000$. Appendix C reports results when I use a different value of M. Intuitively, as long as M is sufficiently high such that the observed number of mills is below M in any of the baseline or counterfactual scenarios, the exact value of M is not important.

[20]As a sensitivity analysis, I estimate the parameters using a subset of size $70,000$, which yields very similar results. Appendix C reports the results.

data and the number of mills generated in the model. The linear fit without an intercept has a coefficient of 1.02. Visual representation of the goodness of fit can be seen in Figure 7(b) and 7(c), which plot the number of mills in each location. The model does a relatively good job capturing the variation in the number of mills across Thailand.

Table 8: Estimated Parameters from NFP

	Point estimate	SE
λ	2.913	0.102
constant	1.790	0.047
I(east)	0.426	0.021
I(north)	-0.319	0.027
I(northeast)	0.326	0.022
I(south)	1.106	0.055
I(west)	0.180	0.032
ruggedness	0.772	0.017
% in season	-0.002	0.001
population density	0.005	0.001

Notes: This table reports the point estimates and bootstrapped standard errors for the scale parameter, λ, and entry cost parameters using the nested fixed-point algorithm. Entry cost is assumed to be a function of the location's region, the ruggedness of the location, the percentage of rice that is grown in season at the location, and the population density. Parameters are estimated using data from 2012-2018.

5.3 Farmers' Production Function

To calculate the change in the production of rice in response to a change in farmer prices, I only need to calibrate the share of intermediate input β. I calibrate the share of intermediate input using data from the socio-economic and labor survey of agricultural households, obtaining $\beta = 0.25$. Details of how I calculate changes in rice production are given in Appendix D.1.

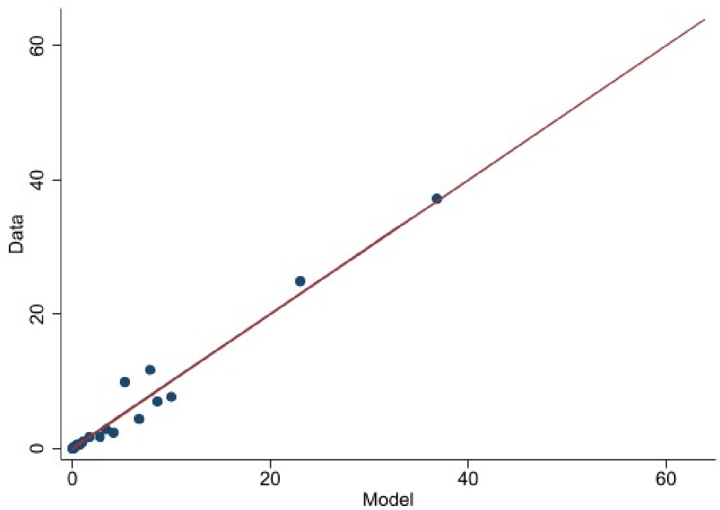
6 Counterfactual Analysis

Having estimated the relevant parameters, I will now examine two counterfactual scenarios, both guided by real-world policies in Thailand. In the first scenario, I study the impact of a country-wide improvement in road infrastructure. In the second scenario, I examine how intermediaries impact farmers' decisions to invest in new technology. To separately examine the role of entry-location decisions of intermediaries, I conduct two exercises for each counterfactual scenario. First, I keep the number of mills in each location the same as in the baseline scenario. Then, I allow mills to change their entry-location decisions.

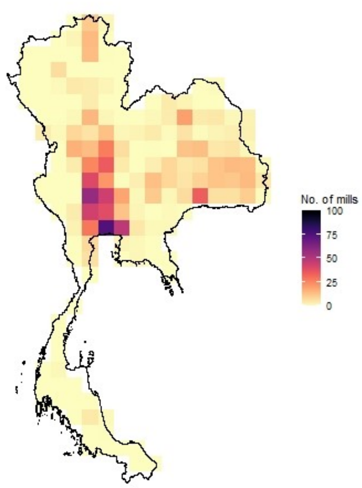
Computation of equilibrium in the baseline and counterfactual scenarios requires data on retail

Figure 7: NFP Goodness of Fit

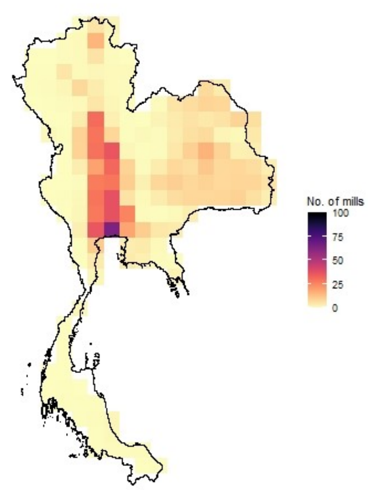
(a) Binned Scatter Plot of No. of Mills



(b) No. of Mills in Data



(c) No. of Mills in Generated from Model



Notes: Number of mills in each location between 2012-2018. Panel (a) shows a binned scatter plot of the number of mills in each location observed in the data against the number of mills generated from the model. The red line is the 45-degree line. Panels (b) and (c) show a heatmap of the average number of mills in each location in the data and the simulated data, respectively. Mill locations consist of 0.5 degree $\times$ 0.5 degree grid cells.

prices and farmers' output, which I will use to deduce farmer's output in the model. I use the average retail prices for white rice and the output per district between 2012-2018 to compute equilibrium in the baseline scenario. Note that since farmer prices and the output depend on the realization of the number of mills in each location, for each scenario, I compute the equilibrium farmer prices and output across 1,000 simulations and average the farmer prices and output across the simulations. In each counterfactual scenario, I keep the retail price the same as in the baseline scenario.[21]

6.1 Improvement in Road Infrastructure

In 2017, the Department of Highways released a strategic plan, setting a goal to reduce highways-related transportation cost down by 9.09%, from 4.4% of GDP to 4% of GDP. Therefore, I map this into the model as a 9.09% country-wide reduction in the iceberg trade costs.[22] To isolate the role of intermediaries' entry decisions from the rest of the channels, I first simulate a reduction in trade costs keeping the intermediaries' entry-location decisions the same as in the baseline scenario. Then, I let the rice mills change their entry-location decisions, allowing the last channel to come into play.

Shutting Off Entry Response of Rice Mills

I first study what happens when I reduce trade costs if I shut down the entry response of the mills. I do this by keeping the number of mills in each location, N, the same as the baseline scenario. The results from this exercise are presented in the first column of Table 9. The aggregate income of the farmers increases by 15.79%. More importantly, the gains to farmers are regressive. The second and the third rows of Table 9 show the percentage change in income of the farmers whose incomes in the baseline scenario are in the top and bottom decile. We can see that the top decile farmers benefit from 11% higher percent increase in income relative to the bottom decile farmers.

To further understand the channels through which a reduction in trade costs impacts the farmers' income, Table 10 decomposes the percentage increase in farmers' incomes that arises from a decrease in trade costs into three different channels: 1) the direct reduction in transport cost, 2) the increase in prices that farmers receive, and 3) the changes in farmers' output. The second channel is a result of higher spatial competition among rice mills, which results from higher outside options of the farmers due to lower trade costs, and the opportunity for farmers to re-optimize which mills to sell their rice to. The third channel arises because farmers re-optimize their intermediate input

[21] The decision to keep the retail price unchanged can be justified by the fact that Thailand exports a large proportion of the rice it produces. According to the USDA (2021), the percentage of rice that Thailand exported ranges from 29.7% of the rice it produces in 2020/21 and 54.5% in 2017/18. Furthermore, the total quantity of rice produced in Thailand forms a very small percentage of the global rice production, producing about 1.1% of the rice produced globally in 2020/21 (USDA, 2021).

[22] Specifically, I reduce τ everywhere by 9.09%. Since the quantity of rice reaching the mill is $\frac{y_f}{\tau_{fm}}$, if τ_{fm} falls by 9.09%, farmers are going to arrive at the mill with 10% more rice than before.

Table 9: % Change in Farmer Income following 9.09% Decrease in Trade Costs

	% Change in Income	
	Baseline N	Flexible N
Aggregate	15.79	16.75
Top 10% (avg.)	16.14	17.27
Bottom 10% (avg.)	14.54	13.84
SD	0.82	1.50

Notes: This table reports the percentage change in farmer income following a country-wide 9.09% decrease in iceberg trade costs. The first column reports the results when the number of mills in each location is the same as in the baseline scenario. The second column reports the results when mills are allowed to change their entry and location decisions. The second and thirds rows show the average percentage change in income of the farmers whose incomes in the baseline scenario are in the top and bottom decile.

usage in response to higher prices. Column 1 shows the percentage increase in income that results directly from lower transport costs paid by farmers to transport rice from their farms to the mills. When τ falls by 9.09%, farmers arrive at the mill with 10% more rice than in the baseline scenario. Therefore, if farmers had produced the same quantity of rice and sold it to the same mills at the same prices as in the baseline scenario, all farmers' incomes would have risen by 10%. The direct effect of reduction in trade cost on farmers' income is homogeneous across all farmers.

Table 10: Decomposition of % Change in Income following 9.09% Decrease in Trade Costs

		Baseline N		Flexible N	
	Transport	Price	Output	Price	Output
	(1)	(2)	(3)	(4)	(5)
Aggregate	10.00	1.63	4.16	2.32	4.44
Top 10% (avg.)	10.00	1.92	4.22	2.65	4.62
Bottom 10% (avg.)	10.00	0.65	3.90	0.20	3.64
SD	0.00	0.59	0.22	1.09	0.41

Notes: This table decomposes the total percentage change in farmer income into the percentage change arising from three different channels. Column 1 reports the percentage point increase in income arising directly from the reduction in transport cost. Column 2 reports the additional percentage point increase arising from the increase in prices that farmers receive, holding the number of mills and the farmers' output the same as in the baseline scenario. Column 3 reports the additional percentage point increase coming from changes in farmers' production decisions, holding the number of mills constant at the baseline level. Columns 4 and 5 report the equivalence of columns 2 and 3 when mills can change their entry and location decisions.

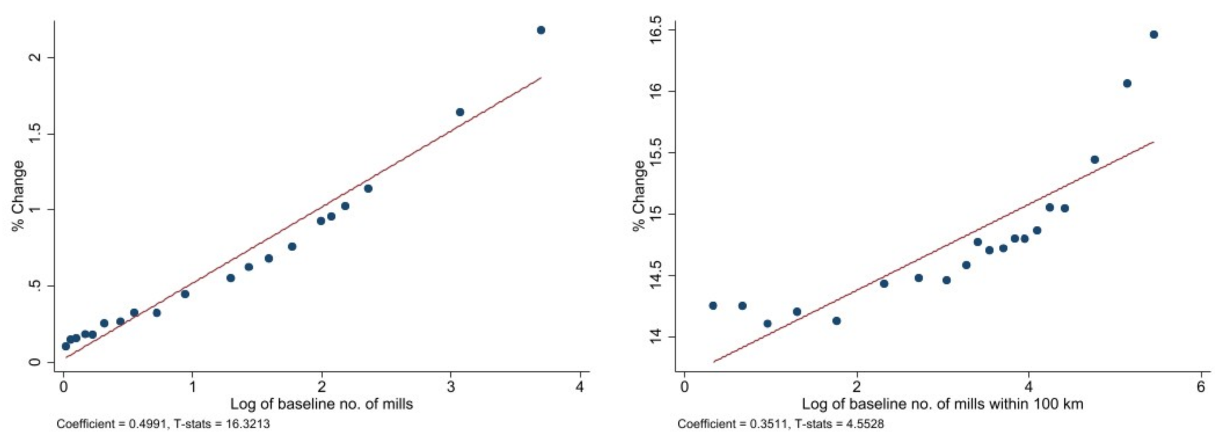
The percentage increase in income arising from further changes in the prices that farmers receive is presented in column 2 of Table 10. Prices that farmers receive increase because of two reasons. First, lower trade costs mean that the value of farmers' threat points in the Nash bargaining process is higher, leading to higher equilibrium farmer prices. Second, farmers reoptimize their choices of mill to trade with. Holding the quantity of rice produced by farmers the same as in the baseline scenario, the aggregate income of the farmers further increase by 1.63 percentage points as a result of changes in prices that farmers receive. Note that the gains arising through this channel are regressive. While the income of top decile farmers further rises by 1.92 percentage points from this channel, the income of the bottom decile farmers only increases by 0.65 percentage points. Likewise, the standard deviation in the percentage increase in farmers' incomes increases from 0 to 0.59 percentage points.

This regressive effect results from the uneven spatial distribution of rice mills in the baseline scenario, which arises from mills' strategic location decisions. Productive places are more profitable; therefore, in the baseline scenario, the number of mills is higher in productive areas. Since farmer prices are determined by equation (19), higher density of mills in productive locations means shorter distances between mills. Therefore, the threat points of productive farmers form a larger proportion of the equilibrium farmer prices. Since lower trade cost impacts farmers' outside option, the reduction in trade costs has larger impacts on farmer prices at productive locations. This is illustrated through the positive correlation between the percentage change in farmer prices and the baseline number of mills in Figure 8(a).

Figure 8: Impacts of Improvement in Road Infrastructure: Baseline N

(a) % Change in farmer prices

(b) % Change in income



Notes: Counterfactual results from 9.09% reduction in the iceberg trade costs when the number of mills in each location, N, is fixed at the baseline level. Panel (a) shows a binned scatter plot of the average percentage increase in farmer prices in each location against the log of the number of mills in the location in the baseline scenario. Panel (b) shows a binned scatter plot of the percentage change in farmer income against farmer access to rice mills, measured using the log of the number of mills within 100 km from the farmer in the baseline scenario.

Subsequent changes in farmers' production decisions further widens the gap between the percentage increase in income of the top and bottom decile farmers. In response to price changes, farmers reoptimize their use of intermediate inputs in their rice production, increasing the quantity of rice that they produce. In particular, for farmer f, the ratio of the new output (y'_f) to the baseline output (y') is a function of the ratio of the new farmer price and iceberg trade cost $\left(\frac{p^{f\prime}_{m'}}{\tau'_{fm'}}\right)$ to the baseline farmer price and iceberg trade cost $\left(\frac{p^f_m}{\tau_{fm}}\right)$:[23]

$$\frac{y'_f}{y_f} = \left(\frac{p^{f\prime}_{m'}/\tau'_{fm'}}{p^f_m/\tau_{fm}}\right)^{\frac{\beta}{1-\beta}} . \tag{33}$$

Column 3 of Table 10 reports the further increase in income once farmers reoptimize their production decisions. Productive farmers who benefit from a larger increase in prices increase their output by more than the unproductive ones, thus widening the gap between the top and bottom decile farmers. The top decile farmers experience a further 4.22 percentage points increase in income after we account for changes in farmers' production decisions. In contrast, the bottom decile only experience an additional 3.90 percentage points increase.

The heterogeneity in the changes in prices that farmers receive and changes in farmers' production decisions lead to the overall regressive effects on the gains to the farmers. Productive farmers, surrounded by a larger number of mills, experience a larger percentage increase in income relative to the unproductive ones. This is reflected in Figure 8(b), which shows the percentage change in farmers' incomes against the number of mills within 100 km from the farmer. Because productive areas have a higher density of mill, mills in productive areas face a greater reduction in market power following a reduction in trade costs, causing productive farmers to experience larger gains.

Allowing for Entry Response of Rice Mills

In this exercise, I incorporate the entry response of rice mills into the counterfactual simulation. The second column of Table 9 shows the percentage change in farmer income once mills are able to change their entry-location decisions. Entry responses of rice mills have important implications on the gains to farmers. The first row of Table 9 shows that after accounting for the entry responses of rice mills, the percentage change in the aggregate income of the farmers increases from 15.79% to 16.75%. Hence, if we had ignored the changes in entry decisions of rice mills, we would have underestimated the percentage increase in aggregate farmer income by 5.7%. More importantly, the entry responses of intermediaries have important distributional consequences. Changes in the entry decisions of the rice mills exacerbate the regressive nature of the gains to the farmer. The gap between the percentage change in income of the top and bottom decile farmers increases by 2.1 times as a result of the entry response of mills.

To shed light on why the intermediaries' entry responses are regressive, let us consider how intermediaries change their entry-location decisions following a reduction in trade costs. Ex-ante, it is unclear how a decrease in trade costs will impact the entry decisions of the rice mills. When trade cost falls, there are three competing forces at play. First, a lower trade cost increases the value of

[23] See Appendix D.1 for details.

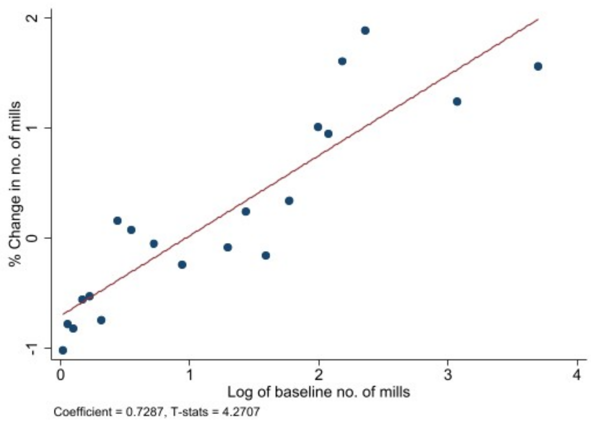
the farmer's threat point in the Nash bargaining problem, putting upward pressure on farmer prices and subsequently lowering per-unit profit, making mills less likely to enter. Second, lower trade costs and higher farmer prices mean farmers increase their output, leading to higher total profit for mills. The increased profitability makes mills more likely to enter. Third, lower trade costs mean that farmers are more likely to travel further to places with higher farmer prices, increasing the profitability in places with high farmer prices and vice versa.

Simulations indicate that the second and the third channels dominate in productive locations, resulting in higher overall profitability. As a result, productive farmers experience a larger increase in income, as shown in Figure 9. In Figure 9(a), we can see the positive correlation between the percentage change in N and baseline N. Since in the baseline scenario, the number of mills positively correlates with the productivity of the location, this means that entry following a decrease in trade costs is regressive; productive locations experience an increase in the number of mills while unproductive locations experience a decrease in the number of mills on average. As a result, locations with higher baseline number of mills experience a larger percentage increase in farmer prices following a decrease in trade costs, as shown in Figure 9(b). This causes a chained response in rice production; farmers in productive locations increase their output by a larger percentage than the unproductive ones, further amplifying the distributional consequences. Figure 9(c) and 9(d) illustrate this regressive effect. After we account for changes in entry-location decisions of rice mills, productive farmers who have higher baseline income and are located in areas with higher mill density experience a larger percentage increase in income relative to when we do not account for entry response of mills. In comparison, unproductive farmers experience a smaller percentage increase in income.

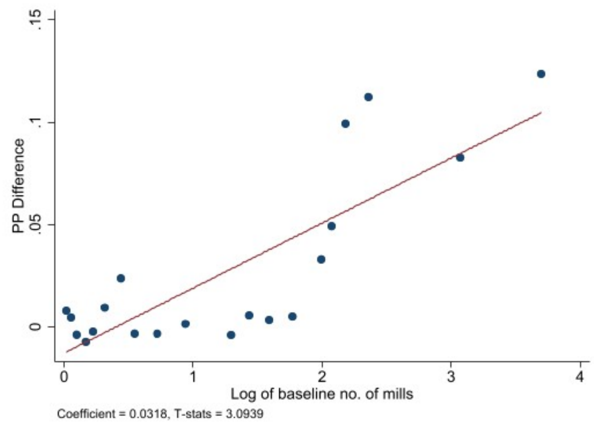
Overall, this counterfactual scenario indicates that intermediaries' entry decisions and market power have important distributional consequences. Entry is regressive. Intermediaries' strategic entry decisions lead to higher density of rice mills in productive areas, resulting in lower market power of rice mills in productive areas. As a result, farmers in productive areas experience a larger percentage increase in income relative to those in unproductive areas. Entry decisions of rice mills following the shock further exacerbate the gap in the gains between the productive and unproductive farmers. Country-wide reduction in trade costs causes productive areas to become relatively more profitable than the unproductive ones, leading to a higher number of mills in productive areas and vice versa. Therefore, entry responses of intermediaries further widen the gap in the percentage increase in farmer prices between the productive and unproductive farmers. Ignoring the entry response of rice mills would cause us to underestimate the gap in the percentage change in income between the top and bottom decile farmers by 53%.

Figure 9: Impacts of Improvement in Road Infrastructure: Flexible N

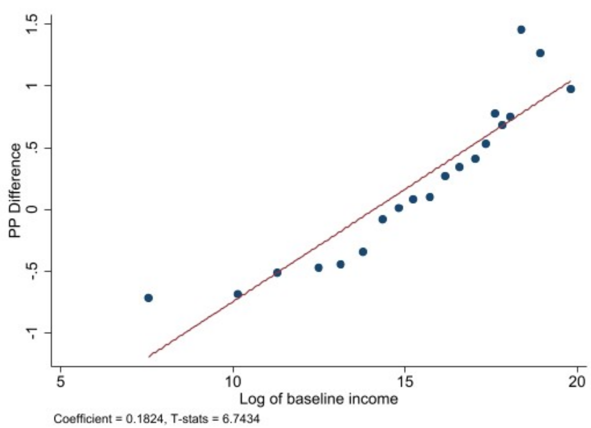
(a) % Change in N vs Baseline N



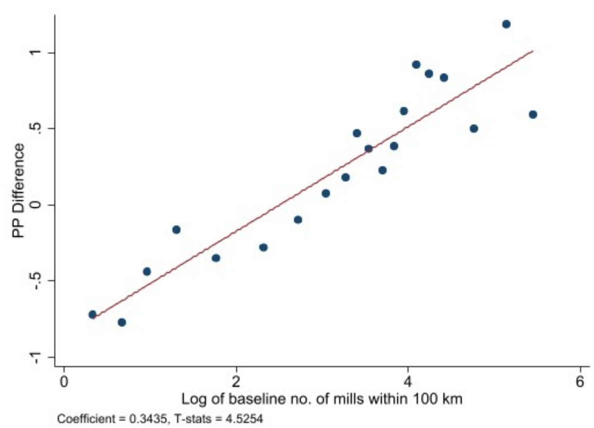
(b) Percentage Point Change in $\%\Delta p_l^f$ after Allowing N to Change vs Baseline N



(c) Percentage Point Change in $\%\Delta$ Income after Allowing N to Change vs Baseline Income



(d) Percentage Point Change in $\%\Delta$ Income After Allowing N to Change vs No. of Mills within 100 km in Baseline Scenario



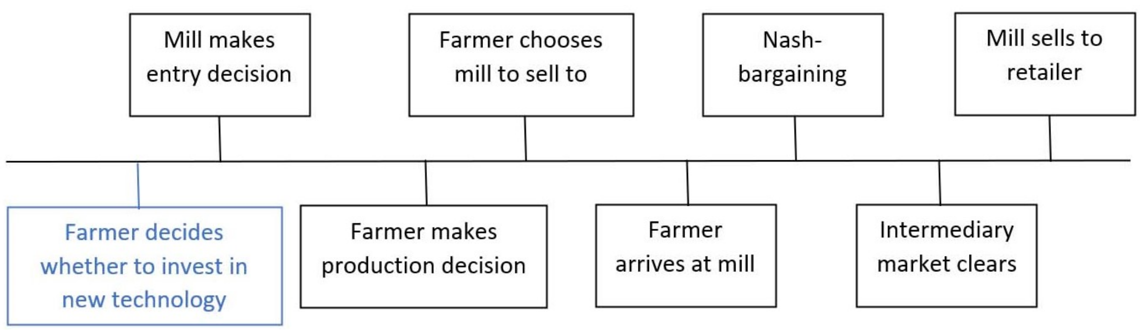
Notes: Counterfactual results from 9.09% reduction in the iceberg trade costs after allowing mills to change their entry-location decisions. Panel (a) shows a binned scatter plot of the percentage change in the number of mills against the baseline number of mills in each location. Panel (b) shows binned scatterplot of the percentage point difference in the percentage change in farmer price before and after the entry response of mills against the baseline number of mills in each location. Panels (c) and (d) show the percentage point difference in the percentage change in farmer income before and after the entry response of mills. A positive percentage point difference means that the percentage increase in farmer income (farmer price) is larger after accounting for mills' entry response.

6.2 Opportunity for Farmers to Invest in New Technology

In the second counterfactual, I consider the role of spatial market power and strategic entry in shaping farmers' decisions to invest in new farming technology. In Thailand, there are many ongoing efforts to encourage farmers to adopt new farming practices or technology. For instance, in 2018, the Thai Rice Department budgeted over 60 million USD for projects aimed to develop the production potential of the agricultural sector (Rice Department, 2018). A significant factor determining the success of these projects is the farmers' uptake of the technology. In what follows, I show that the entry response of intermediaries impacts farmers' decisions to invest in the new technology, resulting in multiple equilibria.

I build this counterfactual scenario on one of the ongoing projects, the Thai Rice NAMA, which encourages farmers to adopt low-carbon emission technology and practices. Mapping this into the model, I allow farmers in the six targeted provinces in the central plains to decide whether to invest in the new technology at the beginning of the period, as shown in Figure 10. The publicly available information on the project suggests that the new technology will increase farmers' productivity by 30% and that farmers will break even at the current prices (NAMA Facility, n.d.). I implement this into the model by setting the investment costs such that farmers will break even at the baseline prices.[24] I assume that farmers will only invest if the return on investment is strictly positive.

Figure 10: Timeline with Opportunity to Invest in New Technology



Because farmers are small, they do not consider how their individual investment decisions impact the aggregate output in the area and subsequently the number of mills in each location. Farmers' investment decisions are determined solely from the expected return from the investment given the expected farmer prices. Since rice mills' density impacts farmer prices and thus the return from investment, farmers' investment decisions are affected by their beliefs on whether other farmers will invest and what the subsequent mill density will be. Multiple equilibria arise, depending on the farmers' beliefs about others' decisions. In the non-socially optimal equilibrium, farmers believe that no one else will invest. They believe that the number of mills and subsequently the farmer prices will be the same as in baseline scenario. Therefore, there will be no positive gains to farmers

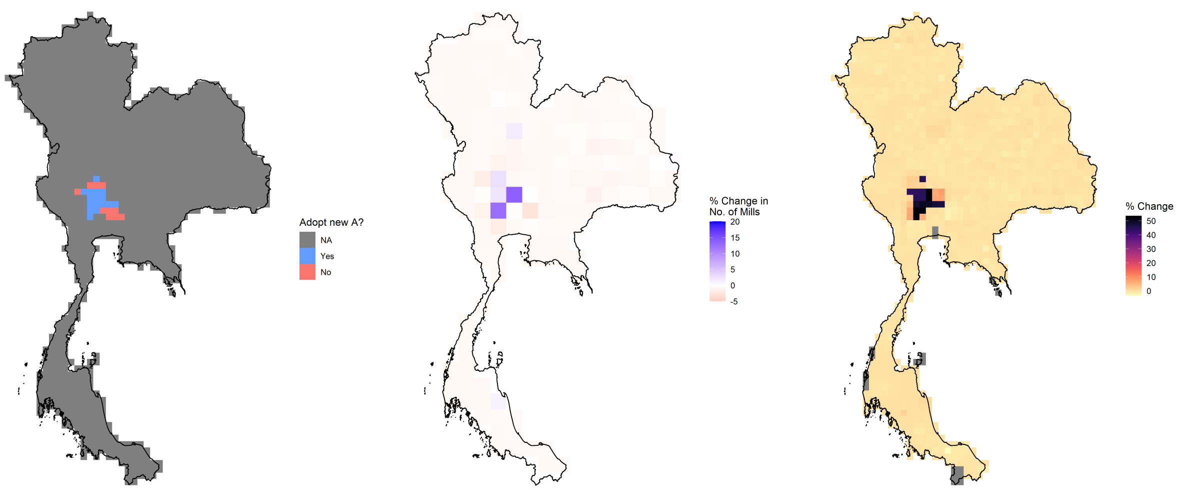
[24] See Appendix D.3 for further details on the project and how I map the project to the model.

from investing in the technology. As a result, no investment is made.[25]

However, such equilibrium is not socially optimal. If farmers collectively invest in the new technology, the increase in output would result in a higher number of mills. Higher number of mills means higher farmer prices. As such, ex-post, the majority of farmers would be willing to invest in new technology. Such equilibrium is shown in Figure 11. If 62% of the targeted farming locations adopt the new technology as shown in Figure 11(a), a higher number of mills will enter those corresponding areas as depicted in Figure 11(b). Subsequently, farmer prices in those areas rise; this increases the return to investment for those farmers such that it is worthwhile for that same 62% of the farming areas to invest in the new technology. As a result, the farmer income rises, as displayed in Figure 11(c).

Figure 11: Technology Adoption in a Socially Optimal Equilibrium

(a) Investment made (b) $\%\Delta N$ (c) $\%\Delta Income$



Notes: This figure shows a socially optimal equilibrium when farmers are presented with an opportunity to invest in new technology. N/A denotes that farmers are not in the provinces targeted by the project and are not offered an opportunity to invest. Panel (a) shows the investment decisions of the farmers. Panel (b) shows the percentage change in the number of mills. Panel (c) shows the percentage change in farmer income.

This counterfactual analysis shows the importance of the entry response of intermediaries in shaping the farmers' investment decisions. Because farmers do not internalize the strategic complementarity arising from endogenous strategic decisions, it is possible to be in an equilibrium in which the level of investment in technology is lower than the socially optimal level.

[25] This equilibrium would also be achieved if farmers are myopic and do not realize that mills' entry decisions are affected by the quantity of rice harvested.

7 Conclusion

The key message in this paper is that strategic entry decisions of intermediaries matter in the presence of trade frictions. Even in the absence of regulatory barriers to entry, due to the existence of fixed costs of entry and trade costs, intermediaries strategically choose their locations. This results in an uneven spatial distribution of intermediaries, with higher density of intermediaries in productive areas. The presence of trade costs means that intermediaries situated in areas with low density of intermediaries enjoy higher spatial market power. Therefore, farmers situated in productive areas, surrounded by a larger number of intermediaries, receive higher prices relative to those in unproductive areas.

While a country-wide infrastructure policy, which reduces trade costs, increases the income of the farmers, the gains from the policy are regressive. Because intermediaries in productive areas have lower market power, farmers in these areas benefit from higher percentage increase in prices relative to those in unproductive areas. The entry response of intermediaries following the policy further exacerbate this regressive effect. Counterfactual simulation indicates that in Thailand, the percentage increase in income of the top decile farmers is on average 25% larger than that of the bottom decile farmers. Ignoring changes in entry decisions of rice mills would lead us to underestimate the gap in gains to the top and bottom decile by 53%.

In addition, the strategic entry decisions and the market power of intermediaries have important implications for farmers' technology adoption. Since farmers may not internalize the strategic complementarity between farmers' adoption of technology and the endogenous entry decisions of intermediaries, farmers can be stuck in a non-socially optimal equilibrium in which there is under-investment in new technology relative to the socially optimal level. I demonstrate the existence of such equilibrium in the context of Thailand. To encourage farmers to adopt new technology, it may not be sufficient for policymakers to simply present the new technology know-how to farmers and expect farmers to invest in the new technology themselves. Since farmers' investment decisions are influenced by the behaviors of their peers, policymakers may need to provide additional incentives, for instance in the form of subsidies, to overcome the initial inertia among the farmers.

References

Aguirregabiria, Victor and Gustavo Vicentini, "Dynamic Spatial Competition Between Multi-Store Retailers," *The Journal of Industrial Economics*, 2016, *64* (4), 710–754.

_ **and Junichi Suzuki**, "Handbook on the Economics of Retailing and Distribution," in Emek Basker, ed., *Emek Basker, ed.*, Edward Elgar Publishing, 2016.

Ahn, JaeBin, Amit K Khandelwal, and Shang-Jin Wei, "The role of intermediaries in facilitating trade," *Journal of International Economics*, 2011, *84* (1), 73–85.

Aker, Jenny C, "Information from Markets Near and Far: Mobile Phones and Agricultural Markets in Niger," *American Economic Journal: Applied Economics*, 2010, *2* (3), 46–59.

Antras, Pol and Arnaud Costinot, "Intermediated Trade," *The Quarterly Journal of Economics*, 2011, *126* (3), 1319–1374.

_ **, Teresa C Fort, and Felix Tintelnot**, "The Margins of Global Sourcing: Theory and Evidence from US Firms," *American Economic Review*, 2017, *107* (9), 2514–64.

Arkolakis, Costas, Arnaud Costinot, Dave Donaldson, and Andrés Rodríguez-Clare, "The Elusive Pro-Competitive Effects of Trade," *The Review of Economic Studies*, 2019, *86* (1), 46–80.

Atkin, David and Dave Donaldson, "Who's Getting Globalized? The Size and Implications of Intra-national Trade Costs," Technical Report, National Bureau of Economic Research 2015.

Attavanich, Witsanu, Sommarat Chantarat, Jirath Chenphuengpawn, Phumsith Mahasuweerachai, Kannika Thampanishvong et al., "Farms, Farmers and Farming: A Perspective through Data and Behavioral Insights," Technical Report, Puey Ungphakorn Institute for Economic Research 2019.

Badinger, Harald, "Has the EU's Single Market Programme Fostered Competition? Testing for a Decrease in Mark-up Ratios in EU Industries," *Oxford Bulletin of Economics and statistics*, 2007, *69* (4), 497–519.

Bergquist, Lauren Falcao and Michael Dinerstein, "Competition and Entry in Agricultural Markets: Experimental Evidence from Kenya," *American Economic Review*, 2020, *110* (12), 3705–47.

Bresnahan, Timothy F and Peter C Reiss, "Entry and Competition in Concentrated Markets," *Journal of political economy*, 1991, *99* (5), 977–1009.

Casaburi, Lorenzo and Tristan Reed, "Using individual-level randomized treatment to learn about market structure," 2021.

Chatterjee, Shoumitro, "Market Power and Spatial Competition in Rural India," 2020. Manuscript submitted for publication.

Chau, Nancy H, Hideaki Goto, and Ravi Kanbur, "Middlemen, Non-Profits and Poverty," 2009.

Conley, Timothy G, "GMM Estimation with Cross Sectional Dependence," *Journal of econometrics*, 1999, *92* (1), 1–45.

Costinot, Arnaud and Dave Donaldson, "How Large Are the Gains from Economic Integration? Theory and Evidence from US Agriculture, 1880-1997," Technical Report, National Bureau of Economic Research 2016.

_ , _ , and Cory Smith, "Evolving Comparative Advantage and the Impact of Climate Change in Agricultural Markets: Evidence from 1.7 Million Fields around the World," *Journal of Political Economy*, 2016, *124* (1), 205–248.

Datta, Sumon and K Sudhir, "Does reducing spatial differentiation increase product differentiation? Effects of zoning on retail entry and format variety," *Quantitative Marketing and Economics*, 2013, *11* (1), 83–116.

de Janvry, Alain, Elisabeth Sadoulet, Dar Manzoor, and Emerick Kyle, "The Agricultural Technology Adoption Puzzle: What Can We Learn from Field Experiments," *Development*, 2016, *178.*

Department of Enviromental Quality Promotion, `https://datacenter.deqp.go.th/service-portal/tsm-network-system/report/member-location-all/`. Accessed September 13,2021.

Department of Highways, "Department of Highways Strategic Plan 2017-2021," `http://www.doh.go.th/uploads/tinymce/goverment/DOH_strategic/Strategic%20Planning.pdf` 2017. Accessed September 13,2021.

Dhingra, Swati and Silvana Tenreyro, "Piggy-Back Exporting, Intermediation, and the Distributional Gains from Trade in Agricultural Markets," 2017.

Dingel, Jonathan I and Felix Tintelnot, "Spatial Economics for Granular Settings," Technical Report 2021. Manuscript submitted for publication.

Donaldson, Dave and Richard Hornbeck, "Railroads and American Economic Growth: A "Market Access" Approach," *The Quarterly Journal of Economics*, 2016, *131* (2), 799–858.

Eaton, Jonathan and Samuel Kortum, "Technology, Geography, and Trade," *Econometrica*, 2002, *70* (5), 1741–1779.

_ , _ , and Francis Kramarz, "An Anatomy of International Trade: Evidence from French Firms," *Econometrica*, 2011, *79* (5), 1453–1498.

— , **Samuel S Kortum, and Sebastian Sotelo**, "International trade: Linking micro and macro," Technical Report, National bureau of economic research 2012.

Edmond, Chris, Virgiliu Midrigan, and Daniel Yi Xu, "Competition, Markups, and the Gains from International Trade," *American Economic Review*, 2015, *105* (10), 3183–3221.

Faber, Benjamin, "Trade Integration, Market Size, and Industrialization: Evidence from China's National Trunk Highway System," *Review of Economic Studies*, 2014, *81* (3), 1046–1070.

Feenstra, Robert C and David E Weinstein, "Globalization, Markups, and US welfare," *Journal of Political Economy*, 2017, *125* (4), 1040–1074.

Gaubert, Cecile and Oleg Itskhoki, "Granular Comparative Advantage," *Journal of Political Economy*, 2021, *129* (3), 871–939.

— , — , **and Maximilian Vogler**, "Government Policies in a Granular Global Economy," *Journal of Monetary Economics*, 2021.

Gourieroux, Christian, Alain Monfort, and Eric Renault, "Indirect inference," *Journal of Applied Econometrics*, 1993, *8* (S1), S85–S118.

Grant, Matthew and Meredith Startz, "Cutting Out the Middleman: The Structure of Chains of Intermediation," 2021.

Hoang, Trang, "The Dynamics of Global Sourcing," Technical Report, mimeo Vanderbilt University 2020.

Holmes, Thomas J and John J Stevens, "Spatial Distribution of Economic Activities in North America," in "Handbook of regional and urban economics," Vol. 4, Elsevier, 2004, pp. 2797–2843.

Iiasa, Fao, "Global agro-ecological zones (gaez v3. 0)," *IIASA, Laxenburg, Austria and FAO, Rome, Italy*, 2012.

Jia, Panle, "What Happens When Wal-Mart Comes to Town: An Empirical Analysis of the Discount Retailing Industry," *Econometrica*, 2008, *76* (6), 1263–1316.

Krusell, Per and Anthony A Smith Jr, "Income and wealth Heterogeneity in the Macroeconomy," *Journal of political Economy*, 1998, *106* (5), 867–896.

Loecker, Jan De, Jan Eeckhout, and Gabriel Unger, "The Rise of Market Power and the Macroeconomic Implications," *The Quarterly Journal of Economics*, 2020, *135* (2), 561–644.

Lowder, Sarah K, Jakob Skoet, and Saumya Singh, *What Do We Really Know about the Number and Distribution of Farms and Family Farms Worldwide?: Background Paper for The State of Food and Agriculture 2014*, Food and agriculture organization of the United Nations, 2014.

Macchiavello, Rocco and Ameet Morjaria, "Competition and Relational Contracts in the Rwanda Coffee Chain," *The Quarterly Journal of Economics*, 2021, *136* (2), 1089–1143.

Matsuyama, Kiminori, "Geographical advantage: Home market effect in a multi-region world," *Research in Economics*, 2017, *71* (4), 740–758.

Melitz, Marc J and Gianmarco IP Ottaviano, "Market Size, Trade, and Productivity," *The Review of Economic Studies*, 2008, *75* (1), 295–316.

NAMA Facility, "Thai Rice NAMA," `http://www.stories.nama-facility.org/better-rice/story/`. Accessed October 13,2021.

Nunn, Nathan and Diego Puga, "Ruggedness: The Blessing of Bad Geography in Africa," *Review of Economics and Statistics*, 2012, *94* (1), 20–36.

Office of Agricultural Economics, "Land Use in Agriculture," `http://www.oae.go.th/assets/portals/1/files/socio/LandUtilization2562.pdf` 2019. Accessed April 23,2021.

Rice Department, "Rice Department 2018 Action Plan," `http://www.oic.go.th/FILEWEB/CABINFOCENTER2/DRAWER050/GENERAL/DATA0003/00003789.PDF` 2018. Accessed October 13,2021.

Rogers, Richard T and Richard J Sexton, "Assessing the Importance of Oligopsony Power in Agricultural Markets," *American Journal of Agricultural Economics*, 1994, *76* (5), 1143–1150.

Rosenstein-Rodan, Paul N, "Problems of Industrialisation of Eastern and South-Eastern Europe," *The Economic Journal*, 1943, *53* (210/211), 202–211.

Seim, Katja, "An Empirical Model of Firm Entry with Endogenous Product-Type Choices," *The RAND Journal of Economics*, 2006, *37* (3), 619–640.

SNRD Asia and the Pacific, "Introducing Thai Rice NAMA," `https://snrd-asia.org/introducing-thai-rice-nama/` March 2017. Accessed October 13,2021.

Sotelo, Sebastian, "Domestic Trade Frictions and Agriculture," *Journal of Political Economy*, 2020, *128* (7), 2690–2738.

Tomar, Shekhar, "Gains from Agricultural Market Reform: Role and Size of Intermediaries," *Available at SSRN 2912434*, 2016.

USDA, "USDA Grain: World Markets and Trade," `https://apps.fas.usda.gov/psdonline/circulars/grain.pdf` April 2021. Accessed September 13,2021.

World Bank, World Development Indicators, "Employment in agriculture (% of total employment)(modeled ILO estimate)," `https://data.worldbank.org/indicator/SL.AGR.EMPL.ZS` January 2021. Accessed September 13,2021.

Zhu, Ting and Vishal Singh, "Spatial competition with endogenous location choices: An application to discount retailing," *QME*, 2009, *7* (1), 1–35.

Appendices

A Empirics: Robustness

A.1 OLS

In this section, I present robustness checks for the OLS regression results. To further address concerns about location-specific unobserved heterogeneity, I use the average province-level farmer prices between 2016-2018 as an additional control variable. These are the years that do not contain data on the month-of-highest-sales and are therefore excluded from the regressions. I average the prices at the province level because the districts and sub-districts in the sample do not always overlap across the years. Table A.1 reports the results. For ease of comparison, the first column contains the results from the baseline specification, which are taken from column 2 of Table 2. Column 2 presents results once I control for the average farmer prices between 2016-2018.

Table A.1: OLS Robustness Checks

	log(farmer price)			
	$r = 100$ km		$r = 75$ km	$r = 125$ km
	(1)	(2)	(3)	(4)
COMP (std)	0.023***	0.019**	0.019***	0.024***
	(0.008)	(0.008)	(0.007)	(0.008)
log(crop sold)	-0.006	-0.003	-0.006	-0.006
	(0.005)	(0.004)	(0.005)	(0.005)
log(province-level output)	-0.012	0.001	-0.012	-0.012
	(0.008)	(0.008)	(0.008)	(0.008)
Crop-suitability-index	0.010*	0.006	0.011*	0.010*
	(0.006)	(0.005)	(0.006)	(0.006)
log(distance to BKK)	0.024	0.016	0.020	0.026
	(0.017)	(0.018)	(0.016)	(0.017)
Average farmers' price in 2016-2018		0.025***		
		(0.007)		
Year FE	Yes	Yes	Yes	Yes
Month-of-highest-sales FE	Yes	Yes	Yes	Yes
Region FE	Yes	Yes	Yes	Yes
Crop-type FE	Yes	Yes	Yes	Yes
N	54,252	54,252	54,252	54,252
R^2	0.511	0.520	0.511	0.511

Notes: This table reports the OLS estimates of equation (2). COMP (std) is the standardized competition measure constructed using equation (1). r specifies the distance cutoff that is used to construct the competition measure. All regressions use farmer prices between 2008-2015 as the dependent variable. Standard errors are adjusted to allow for spatial clustering as in Conley (1999), with a bandwith of 1.5 degree using Bartlett kernel. *$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Additionally, I check for the sensitivity of the competition measure to different distance cutoffs.

I construct competition measures analogous to equation (1) using 75 km and 125 km as distance cutoffs instead of the original 100 km cutoff. Column 3 and 4 of Table A.1 report the results. The positive and significant relationship between local competition level and farmer prices is robust to different distance cutoffs in the competition measure.

A.2 IV

I check the validity of the instrument in several ways. First, I test the sensitivity of the instrument to the distance cutoffs. I use alternative distance cutoffs for the neighbor's productivity and the farmer's own productivity. I define neighbor's productivity as the productivity of land between 75 km - 100 km from the farmer and the farmer's own productivity as the productivity of the land within 25 km from the farmer. Table A.2 reports the results. The results are robust to alternative distance specifications.

Table A.2: IV Results Using Alternative Distance Cutoffs

	COMP (std)	log(price)	No. of mills (std)	log(price)
COMP (std)		0.071***		
		(0.019)		
No. of mills (std)				0.070***
				(0.023)
$Suit^{75-100}$ (std)	0.285***		0.214***	
	(0.036)		(0.050)	
$Suit^{75-100} \times Irri^{75-100}$ (std)	0.159***		0.082**	
	(0.034)		(0.038)	
$Irri^{75-100}$ (std)	0.244***		0.338***	
	(0.059)		(0.117)	
$Suit^{0-25}$ (std)	0.097*	0.015**	0.121***	0.013
	(0.055)	(0.007)	(0.034)	(0.008)
$Suit^{0-25} \times Irri^{0-25}$ (std)	0.060	0.011	0.067**	0.011
	(0.049)	(0.007)	(0.033)	(0.008)
$Irri^{0-25}$ (std)	0.281***	-0.038***	0.181***	-0.030***
	(0.100)	(0.010)	(0.057)	(0.009)
First-stage F-stats	47.952		47.667	
Hansen's p-value		0.862		0.417

Notes: This table reports the IV estimates analogous to those from equations (3) and (4) using alternative distance cutoffs: neighbor's productivity is defined as the productivity between 75 km - 100 km from the farmer and farmer's own productivity is defined as the productivity within 25km from the farmer. COMP (std) is the standardized competition measure constructed using equation (1). No. of mills (std) is the standardized number of mills within 100 km from the farmer. All regressions include controls for the log of quantity of crop sold, province-level output, distance to Bangkok, and fixed effects for the year, the month-of-highest-sales, the region, and the crop-type. All regressions use data from 2008 to 2015. Standard errors are adjusted to allow for spatial clustering as in Conley (1999), with a bandwidth of 1.5 degree using Bartlett kernel. $*p < 0.1$, $** \ p < 0.05$, $*** \ p < 0.01$

Table A.3: IV Results Controlling for Ruggedness

	COMP (std)	log(price)	COMP (std)	log(price)
COMP (std)		**0.087*** **		**0.094*** **
		(0.026)		**(0.025)**
$Suit^{50-100}$ (std)	0.307***		0.301***	
	(0.049)		(0.048)	
$Suit^{50-100} \times Irri^{50-100}$ (std)	0.183***		0.184***	
	(0.037)		(0.037)	
$Irri^{50-100}$ (std)	0.231***		0.229***	
	(0.062)		(0.061)	
$Suit^{0-50}$ (std)	0.070	0.014*	0.067	0.017**
	(0.063)	(0.009)	(0.063)	(0.009)
$Suit^{0-50} \times Irri^{0-50}$ (std)	0.067	0.006	0.067	0.005
	(0.050)	(0.008)	(0.051)	(0.008)
$Irri^{0-50}$ (std)	0.405***	-0.054***	0.404***	-0.055***
	(0.123)	(0.018)	(0.123)	(0.016)
$Rugged^{0-100}$			-0.026	0.034
			(0.057)	(0.022)
$Rugged^{50-100}$	(0.004)	0.018		
	(0.061)	(0.019)		
First-stage F-stats	19.123		20.643	
Hansen's p-value		0.898		0.876

Notes: This table reports the IV estimates analogous to those from equations (3) and (4). COMP (std) is the standardized competition measure constructed using equation (1). No. of mills (std) is the standardized number of mills within 100 km from the farmer. All regressions include controls for log of quantity of crop sold, province-level output, distance to Bangkok, and fixed effects for year, month-of-highest-sales, region, and crop-type. All regressions use data from 2008 to 2015. Standard errors are adjusted to allow for spatial clustering as in Conley (1999), with a bandwidth of 1.5 degree using Bartlett kernel. *$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

There may be a concern that neighbors' productivity may be correlated with provisions of public goods such as roads, which can affect farmer prices. To address this, I use ruggedness as an additional control variable in the IV regression since ruggedness hinders the provision of public goods. Results are reported in Table A.3. Overall, IV estimates are robust to the inclusion of additional control variables.

I further conduct falsification exercises to test the validity of the instrument. I run IV regressions like (3) and (4), but use farmer prices for other crops as the dependent variable instead of the farmer prices for rice. If unobserved factors such as the provision of public goods correlate with the instrument, one would expect the coefficient of the competition measure, instrumented using neighbor's productivity, to be positive and significant. Table A.4 shows that this is not the case.

Table A.4: Falsification Exercises

	Maize	Cassava	Longan	Rubber	Vegetables
COMP (std)	-0.074	0.092	0.102	0.297	-0.136
	(0.080)	(0.093)	(0.344)	(0.196)	(0.177)
N	4,211	5,697	926	1,769	1,209

Notes: This table reports the IV estimates analogous to those from equations (3) and (4). COMP (std) is the standardized competition measure constructed using equation (1). All regressions use the same instrument and control variables as the main IV regression. Standard errors are adjusted to allow for spatial clustering as in Conley (1999), with a bandwidth of 1.5 degree using Bartlett kernel. $*p < 0.1$, $**\ p < 0.05$, $***\ p < 0.01$

B Show that Solution to the Nash Bargaining Problem is a Contraction Mapping

The equilibrium farmer price is given by the a system of equations (19). Define $f : [0, \max_m p^r_m]^{|\mathcal{M}|} \rightarrow [0, \max_m p^r_m]^{|\mathcal{M}|}$ where

$$f(m) = (1-\delta) \max_{k \in \mathcal{M}-\{m\}} \left\{ \frac{f(k)}{\tau_{mk}} \right\} + \delta p^r_m \tag{B.1}$$

and $\tau_{mk} \in [1, \infty]$, $\delta \in (0,1)$, $p^m \in \mathbb{R}$.

The max function is continuous. Therefore, f is a continuous function. By Brouwer's Theorem, a fixed point exists. Chatterjee (2020) shows that f satisfies Blackwell's sufficient condition and is, therefore, a contraction mapping.

C Estimation

C.1 How important are the threat points to farmer prices?

Given the estimated bargaining power parameter, δ, and the trade cost parameter, ϕ, in section 5, how important are the threat points in the Nash bargaining problem? Table the size of the threat point as a percentage of the equilibrium farmer prices. On average, the sizes of farmers' threat points are about 20.13% of the equilibrium farmer prices. However, there is substantial heterogeneity. Farmer's threat points can be as high as 88.62% and as low as 0.54% of the equilibrium farmer prices.

Table C.1: Size of Threat Points

	Threat point as % of farmer price
Average	20.13%
Max	88.62%
Min	0.54%
SD	13.13%

Notes: This table shows the size of farmers' threat points, $\max_{k \in \mathcal{M}-\{m\}} \left\{ \frac{p^f_k}{\tau_{mk}} \right\}$, as a percentage of equilibrium farmer prices. Results are simulated using estimates presented in Table 5.

C.2 Sensitivity Analysis on the Choice of $|S|$ and M

In this section, I check for the sensitivity of the estimated parameters to the specified number of potential entrants, M, and the size of the subset S that I use to approximate the variable profit, as described in section 4.3.1. Table C.2 reports the estimated parameters when I set $M = 3,000$. Table C.3 reports the point estimates for different sizes of S. The alternatives I consider are a subset of size $55,000$, $65,000$, and $70,000$. Overall, the estimated parameters are not sensitive to the choice of $|S|$ and M.

Table C.2: Entry Parameters Estimates Using Different M

	M = 2,000		M = 3,000	
	Point estimate	SE	Point estimate	SE
lambda	2.913	0.102	2.890	0.105
constant	1.790	0.047	2.048	0.048
I(east)	0.426	0.021	0.429	0.021
I(north)	-0.319	0.027	-0.322	0.027
I(northeast)	0.326	0.022	0.327	0.022
I(south)	1.106	0.055	1.114	0.055
I(west)	0.180	0.032	0.182	0.030
ruggedness	0.772	0.017	0.778	0.017
% in season	-0.002	0.001	-0.002	0.001
population density	0.005	0.001	0.005	0.001

Notes: This table reports the point estimates and bootstrapped standard errors for the scale parameter, λ, and entry cost parameters using the nested fixed-point algorithm. Entry cost is assumed to be a function of the location's region, the ruggedness of the location, the percentage of rice that is grown in season at the location, and the population density. Parameters are estimated using data from 2012-2018. $M = 2,000$ is the baseline specification; the estimates in the first two columns are the same as those in Table 8 and are provided here for ease of comparison.

Table C.3: Entry Parameters Using Different $|S|$

	$\lvert S\rvert = 60,000$	$\lvert S\rvert = 55,000$	$\lvert S\rvert = 65,000$	$\lvert S\rvert = 70,000$
λ	2.913	2.686	2.411	2.289
constant	1.790	1.934	2.141	2.278
I(east)	0.426	0.489	0.519	0.529
I(north)	-0.319	-0.323	-0.377	-0.393
I(northeast)	0.326	0.373	0.393	0.425
I(south)	1.106	1.217	1.342	1.410
I(west)	0.180	0.205	0.248	0.259
ruggedness	0.772	0.835	0.930	0.985
% in season	-0.002	-0.002	-0.004	-0.003
population density	0.005	0.005	0.006	0.006

Notes: his table reports the point estimates of the scale parameter, λ, and entry cost parameters using the nested fixed-point algorithm. Entry cost is assumed to be a function of the location's region, the ruggedness of the location, the percentage of rice that is grown in season at the location, and the population density. Parameters are estimated using data from 2012-2018. The first column shows the estimates from the baseline specification; the estimates in the first column are the same as those in Table 8 and are provided here for ease of comparison.

D Counterfactuals

D.1 Calculating Change in Rice Production

In this section, I illustrate how I compute the new output level. Solving the farmer's maximization problem (17) for the first order condition gives the optimal quantity of intermediate input:

$$X = \beta \frac{p^f}{\tau_{fm}} \frac{y}{w^X} \tag{D.1}$$

Plugging this back into the production function gives:

$$y = \left(A\overline{H}^{\gamma}\overline{L}^{\alpha}\right)^{\frac{1}{1-\beta}} \left(\frac{\beta}{w^X}\right)^{\frac{\beta}{1-\beta}} \left(\frac{p^f}{\tau_{fm}}\right)^{\frac{\beta}{1-\beta}} \tag{D.2}$$

Let y' be the new output corresponding to the new price $p^{f\prime}$ and new iceberg trade cost τ'_{fm}, then we have:

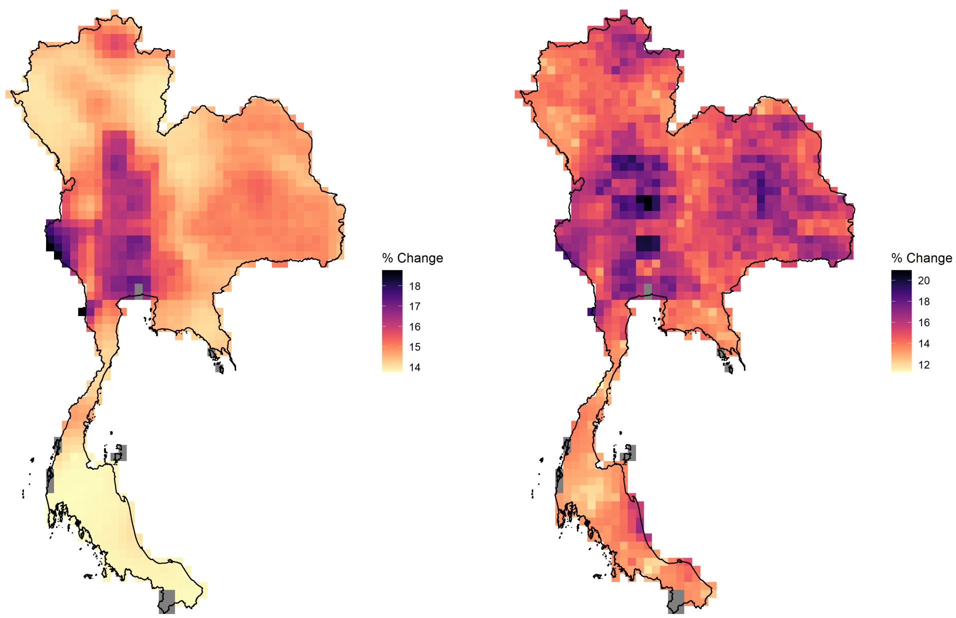
$$\frac{y'}{y} = \left(\frac{p^{f\prime}/\tau'_{fm}}{p^f/\tau_{fm}}\right)^{\frac{\beta}{1-\beta}} \tag{D.3}$$

D.2 Additional Figures for Improvement in Road Infrastructure

Figure D.1: Plot of Percentage Change in Farmer Income

(a) Baseline N

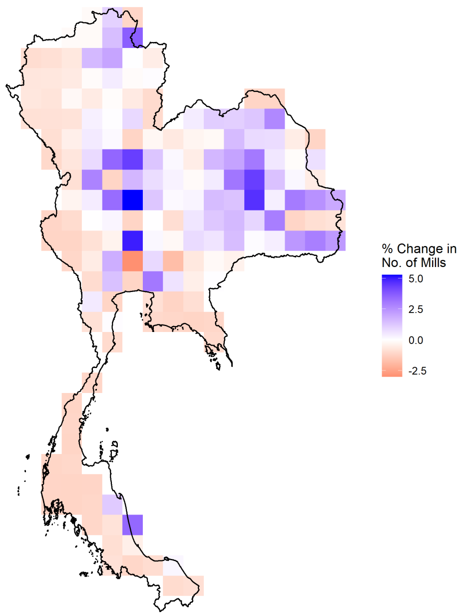
(b) Flexible N



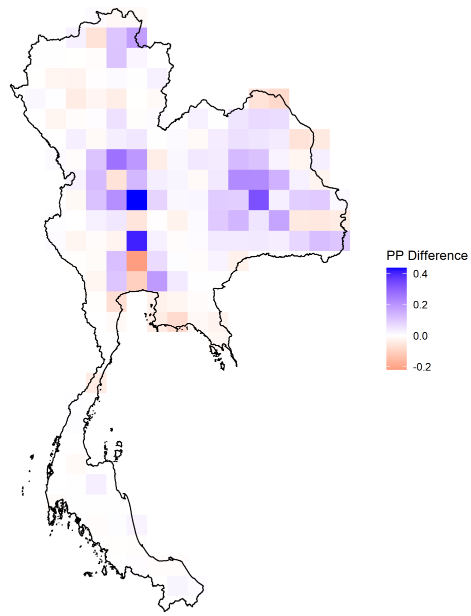
Notes: Percentage change in farmer income after a 9.09% reduction in the iceberg trade costs. Panel (a) shows the percentage change when the number of mills is held fixed at the baseline level. Panel (b) shows the percentage change when mills are allowed to change their entry-location decisions.

Figure D.2: Difference before and after Mills Can Change Their Entry Decisions

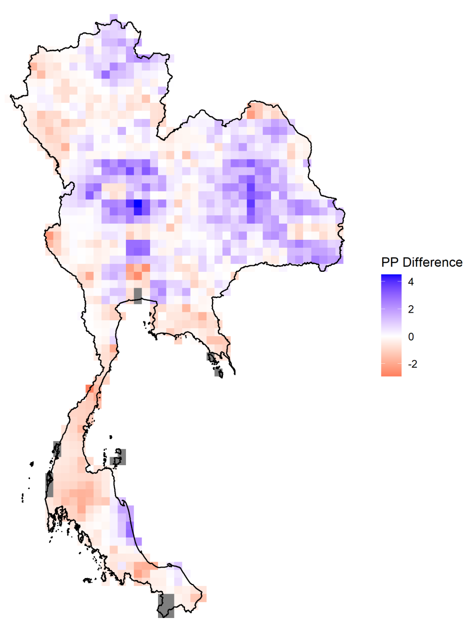
(a) $\%\Delta N$



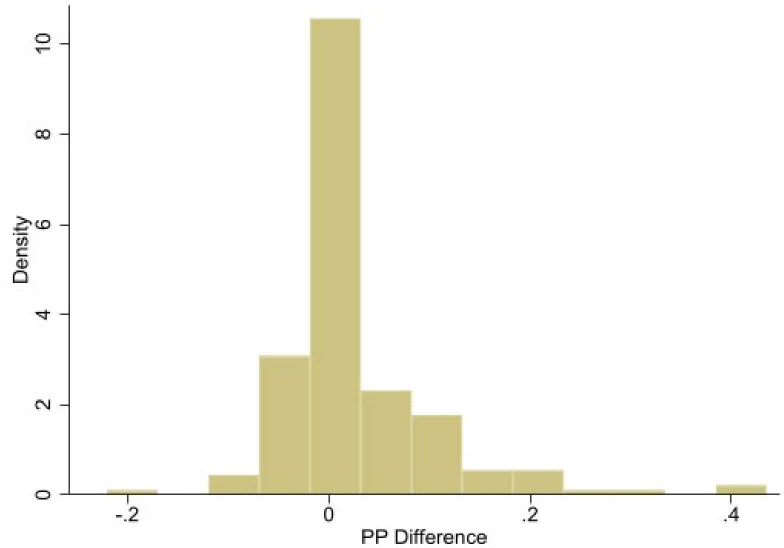
(b) PP Difference in $\%\Delta p_l^f$



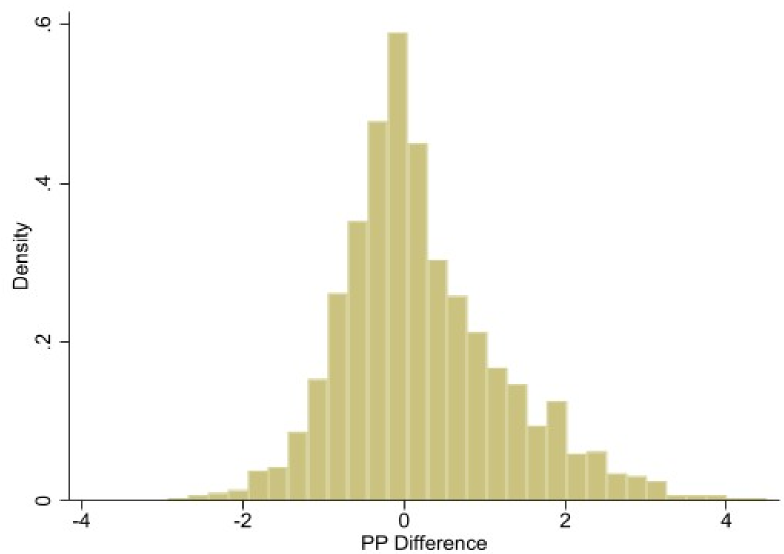
(c) PP Difference in $\%\Delta$ Farmer Income



(d) Percentage Point Change in $\%\Delta p_l^f$ after Allowing N to Change



(e) Percentage Point Change in $\%\Delta$ Income after Allowing N to Change



Notes: Changes in counterfactual results from a 9.09% reduction in the iceberg trade costs before and after mills are allowed to change their entry-location decisions. Panel (a) shows heatmap of the percentage change in the number of mills. Panels (b) and (d) show of the percentage point difference in the percentage change in farmer prices. Panels (c) and (e) show the percentage point difference in the percentage change in farmer income. A positive percentage point difference means that the percentage increase in farmer income (farmer price) is larger after accounting for mills' entry response.

D.3 Thai Rice NAMA

D.3.1 Mapping Projects to the Model

The Thai Rice NAMA aims to promote farmers to adopt low emission technology. The project targets farmers in six provinces in the Central Plains: Chainat, Ang Thong, Pathum Thani, Singburi, Ayutthaya, and Suphanburi (SNRD Asia and the Pacific, 2017). Publicly available information on the project suggests that the technology will reduce farmers' costs by 53% and increase crop yields by 8%. The project expects that the required investment will break even within a year (NAMA Facility, n.d.).

I map the technology improvement into the model, I modify the farmer's production function to take the following form:

$$y_f = A_f \overline{H}_f^{\gamma} \overline{L}_f^{\alpha} (B_f X_f)^{\beta} \tag{D.4}$$

where B is the intermediate input augmenting technology. Using the profit maximizing quantity of intermediate input (D.1), farmer's optimal output is:

$$y_f = \left(A_f \overline{H_f}^{\gamma} \overline{L}_f^{\alpha} \right)^{\frac{1}{1-\beta}} B_f^{\frac{\beta}{1-\beta}} \left(\frac{\beta}{w_f^X} \right)^{\frac{\beta}{1-\beta}} \left(\frac{p_f^f}{\tau_{fm}} \right)^{\frac{\beta}{1-\beta}} \tag{D.5}$$

Let y' be the new output corresponding to the new price $p^{f\prime}$ and new technology A' and B', then we have:

$$\frac{y_f'}{y_f} = \left[\frac{A_f'}{A_f} \left(\frac{B_f'}{B_f} \right)^{\beta} \right]^{\frac{1}{1-\beta}} \left(\frac{p_f^{f\prime}}{p_f^f} \right)^{\frac{\beta}{1-\beta}} \tag{D.6}$$

I map the 8% increase in crop yield as $\frac{A_f'}{A_f} = 1.08$ and the 53% reduction in cost as $\frac{B_f'}{B_f} = \frac{1}{0.47}$. Since $\beta = 0.25$, we have $\frac{A_f'}{A_f} \left(\frac{B_f'}{B_f} \right)^{\beta} = 1.30$, which is equivalent to approximately 30% increase in productivity.

To map the investment cost into the simulation, I assume that the investment will break even in one year and that estimates were made using current prices. I assume that farmers are completely present biased and only care about the return in the current period. Therefore, in the simulation, I assume that the investment cost is such that farmers will break even at the baseline prices. Specifically

$$\text{investment cost}_f = p_{f,baseline}^f \left[\frac{A_f'}{A_f} \left(\frac{B_f'}{B_f} \right)^{\beta} \right]^{\frac{1}{1-\beta}} y_{f,baseline} - p_{f,baseline}^f y_{f,baseline} \tag{D.7}$$

where investment cost_f is the investment cost for farmer f, $p^f_{f,baseline}$ is the price that farmer f receives in the baseline scenario, and $y_{f,baseline}$ is farmer f's output in the baseline scenario.

D.3.2 Algorithm to Compute Socially Optimal Equilibrium

I adopt the following algorithm to compute the socially optimal equilibrium in the second counterfactual scenario.

1. Start by guessing that all targeted farmers invest in the new technology.

2. Compute the Bayesian Nash Equilibrium and simulate the mills' entry decisions.

3. Solve for equilibrium farmer prices.

4. Find which targeted farmers are willing to adopt the new technology at the new equilibrium prices. Farmers will adopt the new technology if

$$\left[\frac{A'_f}{A_f}\left(\frac{B'_f}{B_f}\right)^{\beta}\right]^{\frac{1}{1-\beta}}\left(\frac{p^{f\prime}_f}{p^f_f}\right)^{\frac{\beta}{1-\beta}} y_f p^{f\prime}_f - \left(\frac{p^{f\prime}_f}{p^f_f}\right)^{\frac{\beta}{1-\beta}} y_f p^{f\prime}_f > \text{investment cost}_f \tag{D.8}$$

   where variables without dash denote the baseline level. If farmers who are willing to invest are the same as the guess, then stop. Otherwise repeat step 2-4.

D.4 Perfect Competition

In this section, I consider the counterfactual results if the rice mills have no market power. To compute equilibrium under perfect competition, I assume that there is no markup and that farmer prices are equal to the retail prices.

D.4.1 Improvement in Road Infrastructure

Table D.1 reports the percentage change in farmer income and Table D.2 reports the percentage change in farm gate prices. Note that the percentage change when rice mills have no market power is calculated relative to the baseline where rice mills have no market power. There are two main takeaways from this exercise. First, the gains to farmers are homogeneous when rice mills have no market power. Second, the percentage increase in farmer income is larger when rice mills have market power. The intuition behind the results is as follows.

When rice mills have no market power, the farmer prices, which are the prices that mills give to farmers, are not affected by the trade costs; farmers always receive the retail prices regardless of the trade costs. Therefore, when there is no market power, farmer income only increases because of two channels: 1) direct lower cost of transporting rice from the farm to the mill, and 2) increase in farmers' rice production. Since lower trade costs affects all farmers in the same way, there is no heterogeneity in the gains to farmers.

However, when rice mills have market power, farmer prices are affected by trade costs and rice mills' entry response. Trade costs impact farmers' threat points in the Nash bargaining problem, which impacts the equilibrium farmer prices. In places with higher mill density, farmers' threat points form a larger percentage of the equilibrium farmer prices. Therefore, lower trade costs have greater impact on farmer prices in places with higher mill density. Additionally, mills' entry response affects mill density, which in turn affects farmer prices. Changes in farmer prices generate heterogeneity in the gains to farmers.

Table D.1: % Change in Farmer Income following 9.09% Reduction in Trade Costs

	% Change in Farmer Income		
	No market power	Baseline N	Flexible N
Aggregate	13.55	15.79	16.75
Top 10% (avg.)	13.55	16.14	17.27
Bottom 10% (avg.)	13.55	14.54	13.84
SD	0.01	0.82	1.50

Notes: This table reports the percentage change in farmer income following a country-wide 9.09% decrease in iceberg trade costs. The first column reports the results when rice mills have no market power. The second column reports the results when the number of mills in each location is held the same as in the baseline scenario. The third column reports the results when mills are allowed to change their entry and location decisions. The second and the third columns are the same as the results in Table 9 and are provided here for ease of comparison.

Table D.2: % Change in Farm Gate Price following 9.09% Reduction in Trade Costs

	% Change in Farm Gate Price		
	No market power	Baseline N	Flexible N
Average	10.00	10.88	11.02
Top 10% (avg.)	10.00	11.92	12.65
Bottom 10% (avg.)	10.00	10.65	10.20
SD	0.00	0.59	1.09

Notes: This table reports the percentage change in farm gate prices following a country-wide 9.09% decrease in iceberg trade costs. Farm gate price is the farmer prices accounting for trade cost i.e. $\frac{p_m^f}{\tau_{fm}}$. The first column reports the results when rice mills have no market power. The second column reports the results when the number of mills in each location is held the same as in the baseline scenario. The third column reports the results when mills are allowed to change their entry and location decisions.

D.4.2 Opportunity for Farmers to Invest in New Technology

Under perfect competition, the farmer prices are equal to the retail price and are unaffected by the number of mills. The retail prices are higher that the baseline prices when mills have market power. The return to investment is greater than the investment cost. Therefore, if mills have no market power, there is only one equilibrium in which all targeted farmers invest in the new technology.

E Data Appendix

E.1 Mill

My dataset for rice mills comprise of firms that are registered under rice milling category, specifically under TSIC code 10611. However, firms registered in the dataset may not be actively operating the rice mill in a given year. To address this concern, I only consider mills that report at least 100 THB revenue in their balance sheet in a given year to be operating in that given year. Additionally, to ensure that I do not count firms that are registered under the wrong category in my analysis, I only consider firms that have the words "rice", "mill", or "agriculture" (or the Thai counterpart) in either their name or their listed business purpose as rice mills.

To ensure that firms generally only have one rice mill, I examine the list of rice mills that are active in 2021 from the Department of Internal Trade. In 2021, there are active mills. 95% of these mills are individually owned. In addition, firms that have more than one mill generally have mills in areas with a large number of mills. Therefore, undercounting the mills that belong to those firms do not significantly impact my estimated level of local competition.

Table E.1: Number of Active Mills in 2021 from the Department of Internal Trade

No. of mills per firm	No. of firms
1	1,015
2	18
3	4
4	1