Competition and Relational Contracts: Evidence from Rwanda's Coffee Mills

Rocco Macchiavello — Warwick University

Ameet Morjaria — Harvard University

October 2014*

Abstract

We study the effects of competition (in procurement of raw materials) on relational contracts (RC) using Rwanda coffee mills as a case study. We measure several dimensions of RC between mills and farmers and find i) dispersion across mills in the use of RC, ii) RC practices are correlated with each other, iii) RC are correlated with capacity utilization and unit processing costs. We develop a model highlighting the relationship between competition, RC, mill and farmer outcomes. We estimate an engineering model for the optimal placement of mills to instrument for competition to test the predictions of the model. Competition reduces RC, lowers utilization and increases mill's processing costs. As a result of RC breakdown, we can reject a positive effect of competition on farmers, including increases in prices. The evidence rationalizes policies, such as zoning regulations, monopsony licensing and other entry restrictions, commonly observed in the developing world and emphasizes the importance of promoting contractual enforcement in agricultural value chains.

Keywords: Relational Contracts, Interlinked Transactions, Competition.

JEL Codes: L14, L22, O13, D24, Q13.

---

*_First Draft: Preliminary and Incomplete._ This projects would have not been possible without the collaboration of so many staff members at the Rwanda National Agricultural Exporting Board (NAEB). Without implicating them, we are especially grateful to Alex Kanyankole, Eric Rukwaya, Badrou Bazambanza, Celestin Gatarayiha, Eric Ruganintwali, Jacques Mutabazi, Faustin Nizeymana, Malliavin and Nzamurambaho. We thank David Atkin, Nick Bloom, Oriana Bandiera, Tim Besley, Gharad Bryan, Lorenzo Casaburi, Orie Shelef, Steve Tadelis and Fabian Waldinger as well as conferences and seminar participants at Ascea Summer School, DIME-CEGA (Kigali), ISNIE 2014 (Duke), BREAD Junior Workshop, Gotheborg, Harvard IO, Imperial College, LSE, Nottingham, Oxford and Warwick for helpful comments and suggestions. Financial support from IGC Rwanda Country Program and Harvard Academy WCFIA is gratefully acknowledged.

Corresponding author: Morjaria, Harvard University - WCFIA, email: `amorjaria@wcfia.harvard.edu`.

1 Introduction

When contracts are hard to enforce, parties rely on relationships, or "relational contracts", to sustain trade (Greif (1989, 1993), MacLeod (2007)). Building and managing relational contracts is increasingly seen as a key aspect of management (see, e.g., Gibbons and Henderson (2012)) and a potential source of observed productivity dispersion across firms within narrowly defined sectors (Syverson (2011)). Indeed, greater dispersion in management practices (Bloom et al. (2012)) and productivity (Hsieh and Klenow (2009)) is observed in developing countries, where contracts are harder to enforce. Building and sustaining relational contracts, however, requires rents (MacLeod and Malcomson (1989), Baker et al. (2002), Levin (2003)). This introduces a tension between the positive effects of competition on management (Bloom et al. (2010), Bloom et al. (2014)) and productivity (see, Nickell (1996), Holmes and Schmitz (2010)) highlighted in the literature and the ability to sustain relational contracts (Kranton (1996) and Ghosh and Ray (1996)).[1]

Analyzing the relationship between competition and relational contracts faces two main challenges: the lack of credible measures of appropriate relational practices and the endogeneity of market structure. This paper seeks to address both problems using coffee wet mills in Rwanda as a case study. Besides its intrinsic interest, the context presents a number of advantages.[2] First, multiple contractual imperfections in agricultural value chains in developing countries (see, e.g., Binswanger and Rosenzweig (1982), Bardhan (1989), Fafchamps (2004)) make relationships salient. At the same time, the focus on a single sector with a simple technology allows to measure a number of appropriate relational practices.[3] Second, we take advantage of an engineering model for the optimal placement of mills to instruments for the competition faced by mills in sourcing coffee from farmers. Estimates of the engineering model yield a specific score for the suitability of mill entry for each location (defined at the one square km level) in Rwanda. Controlling for placement suitability and cost drivers within the mill's catchment area, competition is instrumented with scores around the mill's catchment area.[4]

[1] A similar tension has been highlighted in other contexts, e.g., the relationship between competition and innovation (Aghion et al. (2005)) and in the relational banking literature (Petersen and Rajan (1994, 1995)). In both cases rents are required for efficiency.

[2] Coffee mills share many aspects of first stage processing with other agricultural value chains in developing countries. The coffee sector accounts for 30% of Rwandese exports and approximately 15% of its GDP.

[3] We implemented a survey of all mills in Rwanda in 2012. The survey is described below.

[4] The identification strategy combines ideas from two different literatures. GIS data and technological features are used to construct instruments in the evaluation of infrastructure literature (see, e.g.,

Section 2 provides industry background and presents the data. After describing the technology and summary statistics for mills in Rwanda, measures of relational contracting are introduced. Due to lack of inputs, credit, saving and insurance markets, spot market transactions in which mills and farmers simply exchange coffee cherries for cash at harvest time do not guarantee efficiency. Efficient trade requires mills and farmers to establish "interlinked transactions" in which a complex bundle of promises involving services before harvest, at harvest and post harvest is exchanged. Farmers and mills, however, do not have access to legal instruments to enforce these promises. The extent to which trade between mills and farmers is efficient depends on the quality of the relationship between the parties.

We focus on three key aspects of relational contracts: inputs and loans provided by the mill to the farmers before harvest, cherries sold on credit by the farmers to the mill at harvest time, and second payments and assistance from the mill to the farmers post-harvest. We measure each practice from the mill's manager and farmers perspective and aggregate them up into standardized scores. We document three facts. First, there is dispersion in the use of relational contracts across mills. Second, relational practices before, during and post harvest are positively correlated with each other. Third, relational contracts correlate positively with mill's capacity utilization and negatively with processing unit costs. This gives us confidence that the measures of relational contracts are picking up appropriate management practices in the industry.

Section 3 presents a simple theoretical framework. The framework captures the key aspects described in the background section. The model highlights how competition reduces parties *ability* to sustain relational contracting, rather than parties *demand* for it. Whether competition destroys relationships or not depends on parameters. When it does, however, the model delivers a number of clustered predictions: 1) a decrease in the use of *all* practices, 2) a decrease in mill's as well as aggregate capacity utilization, 3) a decrease in either inputs and/or effort by the farmer, and 4) an ambiguous effect on prices and farmer's welfare.

The empirical Sections test the predictions of the model. Section 4 presents the main results on the effects of competition on relational contracts. Competition is defined as the number of mills within a certain distance from each mill. OLS results are presented first. Competition negatively correlates with a broad spectrum of relational practices described above. These correlations, however, could be driven by omitted

Duflo and Pande (2008) and Lipscomb et al. (2013)). Following Bramoulle' et al. (2009) an emerging literature uses neighbors ('s neighbors) to estimate peer effects, of which competition is a special case. Bloom et al. (2014) uses neighbors' local council marginality to instrument for hospital closures in U.K. hospitals.

factors and reverse causality considerations that could bias OLS results in an *a priori* ambiguous direction.

To address these concerns, the IV identification strategy is introduced. First, the four criteria specified in the engineering model are validated in the data. Second, a score for the suitability of mill's placement at the one square km level is obtained. The score (or the individual components of the engineering model and their interactions) is aggregated at the level of the mill's catchment area and surrounding areas to construct the instruments. The exclusion restriction is satisfied if, controlling for placement suitability and cost drivers within the mill's catchment area, suitability for entry around the mill's catchment area affects mill's operation only through its direct impact on competition.

The IV strategy yields a very strong first stage. Second, reduced form results find evidence of a negative relationship between the suitability for entry around the mill's catchment area and the use of relational contracts by the mill. Third, IV results confirm the negative impact of competition on the use of relational contracts. A negative impact of competition is found for essentially all individual relational practices. An additional mill within a 10 kms. radius from the mill reduces the overall relational contract score by 0.1 standard deviations. This effect is equivalent to a jump up or down of fifteen percentiles in the relational contracts ranking for the average mill. The estimates imply that the impact of an additional mill on relational contracts decreases with its distance to the mill and vanishes at approximately ten kms. The IV point estimates are, in absolute value, larger than the OLS estimates, possibly due to omitted factors driving both entry and ability to sustain relational contracts (or to measurement error). Competition negatively impacts measures of the "overall" quality of the relationship between the mills and the farmers, measured through "trust" questions.[5]

When competition leads to the breakdown of relationships, the model predicts a number of additional effects. These are tested in Section 5. The section explores the effects of competition on mill's and farmer's outcomes. As predicted by the model, the breakdown of relationships is associated with worse outcomes at the mill level. Mills suffer lower and more irregular capacity utilization. This leads to an increase in the

[5] We explore the robustness of the IV findings through a number of robustness checks, including: *i*) alternative definitions of catchment areas sizes (both defined exogenously or using mill's specific information), *ii*) alternative strategies to construct the IV (including using individual score components to run over identification tests), *iii*) alternative measures of competition, including aggregate capacity installed and distance to nearby stations.

labour costs to process one unit of output.[6]

The model predicts an ambiguous effect on farmers welfare. If anything, we find a negative impact of competition on farmers' outcomes. Due to lack of appropriate saving technology, farmers process parts of the coffee at home unless they can receive sizeable post-harvest payments from the mills as part of the relational contract. Consistent with this mechanism, competition reduces farmer's share of cherries sold to mills and increases the likelihood of reporting that home processing is used to have cash when needed at the end of the season. Competition instead, doesn't increase prices received by farmers, nor aggregate use of inputs (though farmers pay a higher share) or yields. An overall index of farmers' job satisfaction is reduced by competition. We also detect moderate negative impacts on coffee quality originating from farmer's reduced effort (rather than mill processing). The evidence rejects the hypothesis that competition reduces farmers' demand for the bundle of services exchanged through relational contracts. Finally, section 6 discusses additional "placebo" tests and policy implications.[7]

Besides its relationship with the management literature described above, this paper contributes to the literature on contractual relationships between firms in developing countries.[8] The paper shares with McMillan and Woodruff (1999), Banerjee and Duflo (2000), Macchiavello and Morjaria (2013) and emphasis on relational contracts. McMillan and Woodruff (1999) and Banerjee and Duflo (2000) also rely on cross-sectional survey evidence, but do not consider the importance of competition. Macchiavello and Morjaria (2013) infer the importance of relational contracts exploiting detailed transaction level data and a negative shock in the value chain. Fafchamps (2000, 2004, 2006) has documented the importance of informal relationships between firms in Africa and elsewhere.[9] The paper also contributes to a rich literature on interlinked transactions in agricultural value chains in developing countries. Bardhan (1989), Ray and Sengupta (1991), Mukherjee and Ray (1995), Ghosh et al. (2000) and Kranton and Swamy (2008) provide early theoretical analysis. A nascent empirical

[6] This happens despite aggregate excess supply of cherries in most localities. Higher labour costs arise from Irregular supply of cherries and adjustment costs in the amount of labour employed by the mill and are not due to direct competition for workers.

[7] The cross sectional nature of the identification strategy raises concerns if, holding conditions at the entry site constant, better managed mills strategically entered in areas with less suitable neighboring environments. Evidence from manager's characteristics, the order of entry in the industry, and entry that occurred after our survey in 2013 and 2014 support the validity of our identification strategy.

[8] Gil and Zanarone (2014) provide an excellent review of empirical work on relational contracts.

[9] Banerjee and Munshi (2004), Andrabi et al. (2006), Munshi (2010) and Macchiavello (2010) are examples of studies of contractual relationships in a development context, but with rather different focus.

literature is investigating interlinked transactions involving firms. Blouin and Macchiavello (2013) analyzes interlinked contracts between foreign buyers and exporting coffee mills using loans and contract level data from a specialized international lender. Working in partnership with a dairy cooperative, Casaburi and Macchiavello (2014) conduct a number of experiments to study the implications of farmers' demand for delayed payments on contractual arrangements and market structure. Casaburi and Reed (2013) study interlinked transactions between smaller traders and farmers while Ghani and Reed (2014) study the effect of the entry of additional ice suppliers on the relationships between small ice distributors and fishermen. In both cases, competition appears to be associated with an increase in the amount of credit intermediated along the supply chain.[10]

2 Industry Description

2.1 Industry Background and Data Sources

Coffee Mills

The coffee cherry is the fruit of the coffee plant. The cherries are ripe when they change color from green to red, at which point they should be harvested. The harvest season typically lasts for three to four months. The timing of the harvest season varies by country and, within country, by region depending on altitude, soil and rainfall patterns. Coffee cherries are picked by hand, a labor intensive process that requires significant care and effort. Cherries, even from the same tree, do not get ready for harvest all at once. While less laborious, harvesting cherries all at once compromises quality.

The pulp of the coffee cherry is removed leaving the seed or bean which is then dried to obtain parchment coffee. There are two processing methods to obtain parchment coffee: the dry method and the wet method. In the dry method cherries are cleaned and then dried on tables. This process is done by the farmers at home. In the wet method, instead, cherries are brought at the mill within few hours of harvest. The wet method requires specific equipment and substantial quantities of water. After the cherry skin and some of the pulp are removed with a pressing machine, cherries are then sorted by immersion in water. The bean is then left to ferment, typically for around 30 hours, to remove the remaining skin. The fermentation process has to be

[10] Porto et al. (2011) survey a rich policy oriented literature documenting how episodes of market liberalization have compromised efficiency in export oriented agricultural chains. Little and Watts (1994) offers a review of contract farming from a development studies perspective.

carefully monitored to prevent the coffee from acquiring undesirable flavors. When the fermentation is complete, the coffee is thoroughly washed with clean water. The beans are then spread out on tables and frequently turned by hand until completely and uniformly dry.[11]

The processing method has a significant effect on the flavor of coffee once roasted and brewed. The wet method delivers higher consistency and quality which is reflected in prices. Fully washed coffee is sold at a substantial price premium (around 40%) relative to dry coffee both as parchment and as green coffee at the export gate.[12]

Coffee Mills in Rwanda

Coffee has been an important contributor to the Rwandese economy for several decades. Coffee became widespread in the late 1930s following five waves of mandatory coffee-tree planting imposed by the Belgian colonial administration in an attempt to increase revenue collection from its Rwanda-Urundi colonies (see, e.g., Blouin (2014)). At independence, in 1962, coffee represented 55% of Rwandese exports against minerals (37%), pyrethrum (3%) and tea (2%). Decline in coffee exports started in the mid '80s, accelerated with the demise of the International Coffee Agreement in 1989 (and the subsequent collapse of coffee prices in the global market) and reached its peak with the political instability leading to the 1994 genocide. Since the end of the genocide the sector has steadily recovered. At the time of our survey in 2012, there are around 350,000 farmers, mostly small holders, growing coffee in Rwanda and coffee accounted for almost 30% of Rwandese exports. The number of mills has increased from only one active in 2001 to more than 200 active during the last harvest in 2014 (see Figure 1). It is estimated that the coffee sector accounts for between 12% and 15% of Rwanda's gross domestic product.

The sector still faces challenges. Total installed capacity in 2012 would have allowed the country to process around 70% of the harvested coffee. Export data for coffee harvested in 2012 show that only 30-40% of the exported coffee volumes was washed. In recent years, despite further entry of mills, the percentage of fully washed coffee and the number of operating mills has remained stable.[13]

[11] In other countries, beans are spread out in patios and raked. Drying coffee on tables, the only method used in Rwanda, improves quality but increases cost and labor significantly. After the drying process is completed the coffee is hulled before exports by other downstream firms.

[12] A decomposition of margins along the Rwandese value chain definitively confirms that, despite the higher processing costs, the wet method creates significantly higher value added along the chain. See Macchiavello and Morjaria (2014) for details.

[13] The number of installed (operating) mills has been 199 (190) in 2011, 214 (197) in 2012, 222 (202) in 2013 and 229 (199) in 2014. During this period the volumes and share of fully washed coffee have remained substantially stable.

Data Description

The empirical analysis combines survey data collected through a census of mills with detailed administrative and GIS data collected from a number of different sources (see Appendix for detail). The survey was designed by the authors in collaboration with the National Agricultural Exporting Board (NAEB) and was implemented towards the end of the 2012 harvest season, between June and July 2012. The survey covered all mills operating that season. Each survey team included qualified coffee personnel from NAEB. The mill's manager, the main coffee collector, five randomly selected farmers and four randomly selected workers were interviewed at each mill. The survey covered personal characteristcis from all respondents, the main aspects of each respondent's job and relationship with the mill, and a comprehensive overview of the mill's operations. A detailed capital census and GPS modules were also collected. Finally, random sample of coffee lots were taken from each mill and physically examined and cupped in the coffee board's laboratory in Kigali.[14]

The survey is matched with administrative data obtained from the coffee board and other agencies. There are three main sets of data. First, we assembled a high resolution (1 Km2) GIS database with information on geographic, climatic and infrastructure characteristics for the whole of Rwanda. This is essential to construct environmental controls variables and estimate the engineering model for mill placement. Second, we matched the Rwanda coffee census conducted in 2009 with the GIS data. The census covers all farmers in Rwanda (circa. 350,000) and includes farmers' village location.[15] This provides basic information about trees and production at a highly disaggregated level for places with and without mills. Finally, all mill records are matched to administrative data including weekly volumes and prices for several harvest seasons and transaction-level export records.

2.2 The Operation of Mills

Mill's Main Descriptive Statistics

Given typical firms size distribution in developing countries (see, e.g., Hsieh and Olken (2014)) wet mills are large firms. Table 1 reports summary statistics for mills in Rwanda. The average mill employs a bit more than 80 employees and a supplier base of more than 300 small holders farmers. The employment figures leave mills comfortably

[14] The response rate was nearly 100%. Due to heavy rain it was not possible to reschedule the survey with just one mill.

[15] Rwanda is administratively divided into 4 provinces, 30 districts, 416 sectors, 2148 cells and 14482 villages. The average village is smaller than 2 Km2. This allows us a precise match of the coffee census to the GIS data.

in the right tail of the firm size distribution in Rwanda (see, e.g., Söderbom and Kamarudeen (2013)).[16]

In 2012 the sector in Rwanda was almost equally split between 111 privately owned mills and 105 cooperatives. Relative to cooperative, private mills tend to be larger, have lower utilization rates, higher unit costs and more professional managers.[17]

Mills in Rwanda differ in terms of capacity size. Capacity can be calculated from three aspects of the capital stock invested in the mill (see Appendix for detail): *i*) the number of disks in the pulping machine, *ii*) the metric cubic capacity of the water tanks, and *iii*) the surface of drying tables. Figure 2 reports seasonal capacity estimates for each mill from administrative records.[18] There is large dispersion in installed capacity. The smallest mills have estimated capacity of approximately 100 or 150 tons per season. The bulk of the mills are medium sized with capacity of 500 tons per season. A handful of mills have higher capacity.

"Relational Contracts" in the Industry

We now turn to a description of the use of "relational contracts" in the industry. Given the lack of enforceable contracts and the poor functioning of markets in rural areas (including markets for inputs, extension services, credit, savings and insurance) a well-established literature in development economics has documented the prevalence of interlinked transactions in rural settings (see, e.g., Binswanger and Rosenzweig (1986), Bardhan (1989)). The survey focused on different aspects of these interlinked transactions. We refer to each aspect as a "practice". We distinguish between practices that are relevant pre-harvest season, at harvest season and post-harvest season. For each of these practices, we asked both the farmers and the manager about their use at the mill. We refer to the "relational contract" as the overall set of practices used by the mill and the farmers.

Table 2 presents summary statistics. Before harvest, the mill might have an advantage in providing fertilizers, loans, extension services and other inputs, to the farmers. Approximately 20% of the farmers report to have received fertilizers from the mill and a similar percentage reports to have received loans. While a higher share of managers report to have provided fertilizers to (some) farmers, the figure for loans is similar.

[16] Rwandese mills appear to be somewhat smaller than in other countries. Blouin and Macchiavello (2013) study a sample of 300 mills in 20 countries in which the average mill has yearly sales and assets in excess of $1 million, hires more than 100 employees and source from more than 450 farmers during harvest season.

[17] These differences, however, are largely driven by a number of cooperatives assisted by the NGO TechnoServe that operate smaller mills.

[18] We reconstruct capacity figures from the capital census in the survey and obtain similar figures.

At harvest time, the main aspect of the relational contract is whether cherries are sold on credit to the mill. First, if farmers lack access to a saving technology, they will prefer part of their revenues from coffee to be paid at a later data. Consistently with this hypothesis, a substantial share of farmers reports that one of the two main advantages of home processing is that it allows to sell output when they need the cash. Second, access to working capital credit to finance the purchase of cherries is one of the main operational constraints faced by mills. Purchasing on credit from farmers potentially reduces mill's financial requirements. Approximately 10% of the farmers report to have sold cherries on credit while 30% of the managers report to have purchased some quantity of cherries on credit.

The credit provided by farmers comes in two different ways. First, credit can be very short-term credit, in which case the farmer is paid within a week or so. Second, credit can be implicitly extended to farmers by mills offering a "second payment", typically in the form of a linear bonus depending on volumes sold, at the end of the harvest season.[19] A relatively high proportions of farmers (60% and 70%) expects to receive a second payment at the end of the harvest season. These figures are broadly consistent with what reported by managers: 45% of mills report to have made second payments in the past. Another way in which mills can deliver help to farmers is by helping them in the case of bulky, or unexpected expenses. Among farmers, 64% expect to be able to access help from the mill in case of need while 75% of mills managers report to help farmers with occasional loans for expenses.

Following the predictions of the model, the empirical analysis mainly focuses on the following practices: *i*) before harvest, did the farmer receive inputs from the mill?, *ii*) at harvest, did the farmer sell on credit?, *iii*) post harvest, are there second payments made to the farmer? For each of the three aspects, we focus on both the manager and the farmer answers. After standardizing the responses, we construct scores for the intensity of the relationship before, during and after harvest time by taking the average of the responses given by the manager and the farmers.[20] While we report results OLS and IV results for individual practices and period scores, our main variable of interest is an overall "relational" score which includes both these and additional practices.

There is a certain amount of heterogeneity across mills in the use of relational contracts. The overall relational score ranges from -1.22 to +1.35 and, naturally, greater dispersion is observed in individual scores. The pre-harvest, harvest and post-

[19] Second payments are relatively more common among cooperatives. From a legal point of view members own the coop and might receive a second payment in the form of distributed profits.

[20] That is, we give equal weight to the answer from the manager and the average answer from the farmers.

harvest score are also positively correlated with each other. The correlation is stronger between pre-harvest and post-harvest, and slightly weaker with the harvest score.

(Dispersion in) Unit Costs and Capacity Utilization

Mills are characterized by a relatively simple technology. It takes approximately 5.50 kgs of coffee cherries to produce 1 Kg of parchment coffee.[21] We shall follow industry practices and benchmark the relative efficiency of mills focusing on the costs of producing 1 Kg of parchment (unit costs). The direct costs of purchasing coffee cherries typically accounts for approximately 60% to 70% of unit costs. By working through the stations accounts for the season together with the managers, we obtained accurate measures of unit costs and their breakdown across components. For descriptive purposes we decompose unit costs of mill i as follows:

$$UC_i = \left(P_i^{kg} \times CR_i\right) + OC_i \tag{1}$$

where UC_i are the unit costs, P_i^{kg} is an average price per kilogram of cherries paid by the mill (including estimates for second payments), CR_i is the conversion ratio at the mill and OC_i are other costs, mainly labour, finance, transport and procurement.[22]

Figure 3 and Table 3 document the dispersion in unit costs across mills and its components. Figure 3 shows that a significant proportion of the dispersion in unit costs is explained by differences in geographic characteristics of the area in which the mill operates (suitability for coffee, availability of trees, elevation, slope, etc.) and by installed capacity. After purging the data from differences in costs driven by these factors, Figure 3 still documents significant dispersion in unit costs (the 90/10 percentile ratio is equal to 1.5).

Table 3 decomposes unit costs. As expected, there is no dispersion in the conversion ratio CR_i (90/10 ratio is lower than 1.1). There is more dispersion in the prices paid to farmers (90/10 ratio equal to 1.32). The bulk of the dispersion, therefore, originates in the components of unit costs which are more directly influenced by management: labour, capital, procurement and logistic. Here we find a 90/10 ratio equal to 2.32.

Capacity utilization is an important correlate of unit costs. As a summary statistic, we measure utilization as the total amount (in tons) of cherries processed during the harvest season divided by the total capacity installed.[23] Consistently with Table 3,

[21] The exact conversion ratio depends by coffee variety and other geographical factors affecting the organic properties of coffee.

[22] Labour, transport and procurement costs are easy to compute from the accounts. As usual, capital costs require additional assumptions.

[23] The mill seasonal theoretical capacity is computed assuming a certain length of the season. While

Figure 4 documents significant dispersion in utilization rates during the 2012 harvest season. Leaving aside inactive mills (those that had zero capacity utilization), the median mill has a capacity utilization around 50%.[24]

Figure 5 shows that, conditional on a host of geographic characteristics and mill controls, the relational contracts score, capacity utilization and unit costs are correlated with each other. Prima facie, the measure of relational contracts captures aspects of managerial practices appropriate in this industry.

3 Theory

This section lays out a simple theoretical framework to illustrate the main forces at work in the empirical context we examine. First, we present the set up of the model. Besides describing preferences and technology, we illustrate interlinked transactions between mills and farmers when contracts are perfectly enforceable. We then study the contractual outcomes between the mill and the farmers under different scenarios. First we examine a monopolist mill that can only rely on spot transactions with the farmers. We then explore the rationale for and the conditions under which relational contracts between the monopolist mill and the farmers in the surrounding areas are feasible. We then introduce competition from another mills. While the impact of competition on the mill's costs is shown to be unambiguously negative, the exact mechanisms at work and the final impact on prices paid to farmers depend on parameters. We conclude the section discussing extensions and alternative frameworks.

3.1 Set-Up

A mill operates in an area populated by a unit mass of identical farmers, indexed $i \in [0, 1]$. Each farmer produces a quantity of coffee cherries q. Time is represented by an infinite sequence of seasons, indexed $t = 0, 1, 2..., \infty$. Within each season, there are three subperiods, corresponding to pre-harvest (sub-indexed by 0), harvest (sub-indexed by 1) and post-harvest (sub-indexed by 2). The quantity q becomes available at harvest. Farmers derive utility from consumption at harvest, c_1, and post-harvest, c_2, with preferences given by $\min\{c_1, c_2\}$. These preferences capture in a parsimonious way

we do not have mill downtime, we are computing alternative measures of capacity utilization using weekly tons of cherries purchases from the mills. Preliminary results confirm the simpler analysis reported here.

[24] Capacity under-utilization is not driven by lack of available cherries to process. Highly disaggregated estimates of local production derived from the coffee census suggest that the majority of localities have installed capacity lower than potential local production.

farmers' demand for within-season consumption smoothing.[25] The mill is risk neutral and only cares about expected (discounted) profits. All parties have a discount factor $\delta < 1$ across seasons while, for simplicity, there is no discount within season.

Coffee cherries q must be processed immediately after harvest. Once processed, cherries can be stored up to the following post-harvest period. Two technologies are available: home processing and wet processing. Both technologies yield one unit of output for each unit of cherries. Home processing is performed by the farmer at home and, for simplicity, we assume this process has no additional cost. The farmer can sell home processed coffee at an exogenous price ρ at harvest and post-harvest. Denote by $q_{\rho 1}$ and $q_{\rho 2}$ the quantities the farmer sell on the home processed market at harvest and post-harvest time respectively. Wet processing requires cherries to be given to the mill immediately at harvest time. Let us denote with $q_m(i)$ the quantity sourced by the mill from each farmer i. The aggregate quantity sourced by the mill is then given by $Q = \int_0^1 q_m(i)di$. The mill has en exogenously given installed capacity, $\mathbf{C}$. The mill incurs additional processing unit costs, such as transport, labour, etc., denoted $c(Q) > 0$. The mill is a price taker in the downstream and in other input (e.g., capital, labour) markets. We assume $c'(Q) \leq 0$ provided the mills operates below capacity, $\mathbf{C}$, i.e., $Q \leq \mathbf{C}$, and $c(q) = \infty$ otherwise. Relative to home processing, wet processing increases the value of output in the downstream market. Each unit of output sold by the mill yields $v > \rho$.

Finally, during pre-harvest season, the mill and the farmer can undertake specific investments to enhance production. Specifically, the mill can extend inputs (such as training, fertilizers, loans) to the farmers at cost κ per farmer. If the farmer exerts effort, at personal costs e, to use the inputs provided by the mill correctly, the quality of wet-processed coffee is increased and its value enhanced by a factor λ. We make the following assumption on the parameters:

Assumption 1: $2\frac{\kappa+e}{\lambda q} > v > \max\left\{\rho + c(0), \frac{\kappa+e}{\lambda q}\right\}$.

The assumption $v > \rho + c(0)$ guarantees that wet processing is efficient. The assumption $v > \frac{\kappa+e}{\lambda q}$ guarantees that the mill and the farmers' investments during pre-harvest are efficient. The assumption $2\frac{\kappa+e}{\lambda q} > v$ simplifies exposition by reducing the number of cases to be considered, without substantially altering the results.

[25] Provided preferences display a demand for consumption smoothing within season, the specific functional form of the utility function can be relaxed at the cost of slightly more tedious algebra without altering the main insight of the analysis.

3.2 Benchmark Case: Perfect Contract Enforcement

We begin by considering the case of perfect contract enforcement. When contracts are perfectly enforceable, parties can commit to promises and effort decisions are contractible. For simplicity, let us assume that at the beginning of each pre-harvest season the mill makes a take-it-or-leave-it offer to every single farmer i. Specifically, the mill offers a contract $\mathcal{C}^i = \{\mathbf{I}_k^i, \mathbf{I}_e^i, q_m^i, P_1^i, P_2^i\}$ specifying whether farmer i receives inputs or not, $\mathbf{I}_k^i \in \{0,1\}$, whether farmer i must exert effort, $\mathbf{I}_e^i \in \{0,1\}$, the quantity to be sold by farmer i to the mill at harvest, q_m^i, as well as payments from the mill to farmer i at harvest, P_1^i, and post harvest, P_2^i - the so called "second payment".[26] Each farmer i independently decides whether to accept or reject the contract. If the farmer reject the contracts, she harvests quantity q, process it at home, and sell it on the market at harvest and post-harvest time to maximize her utility. Denote the value of this choice by u^i. If the farmer accepts, all elements of the contract must then be respected by all parties. Denote by $u_{\mathcal{C}}^i$ the utility of farmer i from accepting a contract $\mathcal{C}$. Let us focus on a symmetric solution in which all farmers are offered the same contract $\mathcal{C}^i$. The mill offers the contract that maximizes profits subject to the feasibility constraint and the farmers' participation constraints, i.e.,

$$\begin{aligned} & \max_{\mathbf{I}_k, \mathbf{I}_e, q_m, P_1, P_2} \left((1 + \lambda \mathbf{I}_k \mathbf{I}_e) v - c(q_m) \right) q_m - (P_1 + P_2) - \mathbf{I}_k \kappa \qquad (2) \\ \text{s.t. } u_{\mathcal{C}} \quad & \geq \quad u \text{ and } q_m \leq q. \end{aligned}$$

First, let's derive the farmer's outside option u. A farmer that process at home the entire produce q sells on the market $q_{\rho 1}$ at harvest and stores $q_{\rho 2}$ until post harvest to maximize her utility. Given assumptions, the farmer sells half of her produce at harvest and half at post harvest, i.e., $q_{\rho 1} = q_{\rho 2} = q/2$, obtaining utility $\min\{c_1, c_2\} = \rho q/2$. Second, the farmer's participation constraint must be binding. By assumption 1, i) the feasibility constraint must also be binding at the optimum, i.e., $q_m = q$, and ii) the optimal contract specifies that the mill pays the cost κ, i.e., $\mathbf{I}_k = 1$, and that the farmer exerts effort e, i.e., $\mathbf{I}_e = 1$. Finally, the cheapest way for the mill to satisfy the farmer's participation constraint is to equate farmer's consumption in the harvest and post-harvest seasons. Given $q_m = q$, the firm must set $P_1 = P_2 = P$ such that $P - e = \rho q/2$. In sum,

[26] While the analysis under the benchmark case of perfect contracting is straightforward, an appropriate definition of the contract terms is useful to compare outcomes across scenarios and to match the predictions of the model to the measurement of interlinked transactions in the empirical analysis.

Observation 1: *Under perfect contract enforcement 1) the mill provides inputs to all farmers, 2) all farmers exert effort resulting in high quality, 3) farmers only sell to the mill, 4) the mill pays both a spot price and a second payment, 5) the net present value of the price paid by the mill and the second payment is larger than what farmers obtain in the market, 6) the mill's unit costs of processing cherries are lowest given local growing and operating conditions.*

3.3 Monopoly Mill under Spot Market Transactions

We now turn to the case in which formal contracts between the mill and the farmers are not enforceable. As is well known, when this is the case, the mill and the farmers might rely on a relational contract in order to sustain efficient trade. Before turning to the analysis of the repeated relationship, however, it is instructive to consider the case in which the mill and the farmers only rely on spot transactions, i.e., contracts are enforced only within a subperiod. The timing of events is now as follows. First, the mill sets $\mathbf{I}_k$ deciding whether to pay κ. Then the farmer decides whether to exert effort, $\mathbf{I}_e$. Then harvest is realized. The mill posts a unit price for the harvest season, p_1, and promises an additional second payment, P_2, for the post-harvest season.[27] Given the posted prices and her beliefs, the farmer decides how much to sell to the mill and how much to process at home (and, if any quantity is processed at home, when to sell it). At post-harvest time, the mill decides whether to pay the second transfer promised to the farmer, if any.

The model has to be solved by backward induction. In the absence of a future relationship, the mill always defaults on any payment promised for post-harvest. The farmer, therefore, doesn't believe any promise of second payment and bases her decision entirely on the posted price at harvest time. Suppose the mill has posted a price p_1 at harvest time. Given p_1, farmers decide how much to supply to the mill. The farmer equates consumption at harvest, $c_1 = p_1 q_m + \rho q_{\rho 1}$, with post-harvest consumption, $c_2 = (q - q_m - q_{\rho 1})\,\rho$ subject to the constraint $q_m + q_{\rho 1} + q_{\rho 1} \leq q$. The farmer's supply

[27] We abstract from the potential hold-up problem implied by the perishability of the cherries. Cherries are typically brought to the mill gate or purchased at collection sites, rather than collected at the farmer's gate. The mill collector could hold-up the farmer and renegotiate price downward. Modeling this additional friction would only strengthen the logic of our results. Note also that in this simple environment there is no loss in generality in assuming the mill posts unit prices rather than non-linear price schedules.

curve is then given by

$$q_m(p_1) = \begin{cases} 0 & \text{if } p_m \leq \rho \\ \frac{\rho}{p_1+\rho} q & \text{otherwise} \end{cases} .^{28} \tag{3}$$

The mill, taking as given the supply curve of each farmer, sets p_1 to maximize profits. The mill's problem is given by

$$\begin{aligned} & \max_{p_1} \left((1 + \lambda \mathbf{I}_k \mathbf{I}_e) v - c(Q(p_1)) - p_1 \right) Q(p_1) \\ \text{s.t., } q_m &= q_m(p_1) \text{ for each farmer } i \text{ and } Q(p_1) \leq \mathbf{C}. \end{aligned} \tag{4}$$

Given $c'(\cdot) \leq 0$, the mill optimal solution is to set $p_1 = \rho$ and source as many cherries as possible from each farmer, i.e., $q_m = q/2$. Finally, the price posted by the mill and the farmer's sales decision at harvest are independent of the value of the cherries. As a result, the farmer has no incentive to exert effort, i.e., $\mathbf{I}_e = 0$, and, consequently, the mill offers no input during the pre-harvest season, i.e., $\mathbf{I}_k$. In sum,

Observation 2: *Under spot transactions 1) the mill does not provide inputs to the farmers, 2) farmers do not exert effort to enhance quality, 3) farmers sell only a fraction of their produce to the mill, 4) the mill does not pay any second payment, 5) the net present value of the price paid by the mill and the second payment is larger than what farmers obtain in the market, 6) the mill's unit costs of processing cherries are higher than under full contract enforcement.*

3.4 Relational Contracts with a Monopoly Mill

We now consider (symmetric) relational contracts between the mill and the farmers. When contracts are not enforceable, the relationship between the mill and the farmer faces a two-sided moral hazard problem. First, the farmer must be incentivized to exert effort. Second, the station must be incentivized to keep promises of second payments. A relational contract is a plan $\mathcal{C}^R = \{\mathbf{I}_k^t, \mathbf{I}_e^t, q_m^t, P_1^t, P_2^t\}_{t=0,1,...}^{\infty}$ that specifies investment and effort decisions, $\mathbf{I}_k^t$ and $\mathbf{I}_e^t$,quantities to be delivered at harvest, q_m^t, and payments at harvest and post-harvest, P_1^t and P_2^t, for all future seasons. Parties agree to break

[28]Note that for $p_1 \geq \rho$ the farmer supply curve to the mill is decreasing in the price offered by the mill p_1. The extreme complementarity assumed in the farmers' preferences implies that the income effect always dominates the substitution effect. While this specific property of the farmer's supply curve can be overturned by assuming less extreme demand for savings, the intuition for the main results remains valid provided there is some demand for saving the farmers satisfy by engaging in home processing.

up the relationship and obtain their outside options forever following any deviation. We are interested in deriving conditions under which a stationary relational contract in which *i*) in the pre-harvest season both the mill and the farmer invest, i.e., $\mathbf{I}_k^t = 1$ and $\mathbf{I}_e^t = 1$, *ii*) farmers sell a quantity q_m to the mill, and *iii*) the mill make payments P_1 and P_2 at harvest and post-harvest seasons to each farmer, can be sustained. We assume that if either *a* farmer (by, e.g., selling a quantity lower than q_m^t and/or not exerting effort) or the mill (by, e.g., defaulting on the agreed second payment P_2 to *any* farmer) renege on any contract, both parties revert back to trading using spot transactions forever.[29]

Let us denote the seasonal profits of the mill along the stationary equilibrium by π_r. Denote with π_s the profits of the mill under spot transactions. The incentive compatibility constraint of the mill is given by

$$\frac{\delta}{1-\delta}\left[\pi_r - \pi_s\right] \geq P_2. \tag{5}$$

Denoting with u_r the monetary utility of the farmer along the stationary equilibrium and by u_s the monetary utility of the farmer following a defection, the farmer exerts effort if

$$u_r - u_s \geq e \tag{6}$$

and sell the stipulated quantity to the mill if

$$u_r \geq u_s. \tag{7}$$

The mill offers a relational contract to maximize profits subject to the three incentive constraints above. First, notice that constraint (7) is implied by (6). Second, at the mill's optimum, it has to be that (6) binds. Third, as derived in the previous section, $u_s = \rho q/2$, and, therefore, $\pi_s = \frac{q}{2}\left(v - \rho - c(\frac{q}{2})\right)$. Fourth, the cheapest way for the mill to achieve a certain level of utility for the farmer is to equalize farmer's consumption at each harvest and post-harvest. Along the equilibrium path we must have $P_1 + \rho q_{\rho_1} = P_2 + \rho q_{\rho 2} = \mathbf{P}$. Denoting with q_r the quantity sold to the mill, constraint (6) implies that $\mathbf{P} \geq \rho q/2 + e$. Substituting the constraint (6) into the the mill's incentive compatibility constraint (5), the problem of the mill is to source the

[29] Although the assumption that *all* farmers punish the mill following a deviation against *any* farmer can be relaxed, with a continuum of farmers, individual relational contracts in which punishment is carried out only by the farmer who was cheated on might not exist. Indeed, an important advantage of large organizations is precisely the greater reputation enabled by collective punishment, see, Ghatak et al. (2014) for a model and Casaburi and Macchiavello (2014) for an empirical illustration. Levin (2002) provides an analysis of multilateral relational contracts with a finite number of agents.

maximum quantity q_r subject to

$$\frac{\delta}{1-\delta}\left[(1+\lambda)vq_r - c(q_r) - e - \kappa - (vq/2 - c(q/2))\right] \geq P_2. \tag{8}$$

Note that, for any $q_r < q$ the mill has an incentive to reduce p_2 to a minimum in order to relax the incentive constraint. A necessary condition for this is for the farmer to only sell the residual quantity equal to $q - q_r$ in the post-harvest season. That is, in equilibrium, the farmer never sells to the market during harvest time. The constraint can therefore be rewritten as $P_1 = P_2 + \rho\,(q - q_r)$ and, using (6), we obtain $P_2 = \rho\,(q_r - q/2) + e$. It is then easy to show that, if any relational contract can be sustained at all, it must have $q_r = q$.[30] In sum,

Observation 3: *There exists a critical threshold $\delta_r < 1$ such that if $\delta \geq \delta_r$ a relational contract between the mill and the farmer is sustainable. The relational contract than achieves the efficient outcome and transactions occur as described in the perfect contract enforcement case.*

3.5 Relational Contracts Under Competition

We now consider the case in which a competing mill locates nearby the existing monopolist and can, potentially, try to source cherries from the same farmers at harvest season. We assume this new mill competes only by offering spot transactions in a particular season.[31] As usual, some of the results in this type of models depend on the exact way in which the competition protocol is specified (see, e.g., Tirole (1998)). Rather than focusing on a specific protocol and fully characterize the resulting equilibrium, let us simply focus on deriving conditions under which the threat of competition makes *any* relational contract between the mill and the farmers unsustainable. Denote with $\mathbf{p}_0$ the highest price the competing mill is willing to pay in a spot transaction. If the farmer accepts the offer, she would need to sell $\mathbf{q}_0$ to the competing mill and set $\mathbf{p}_0\mathbf{q}_0 = (q - \mathbf{q}_0)\,\rho$.

Depending on $\mathbf{p}_0$, there are two cases, depending on whether it is more costly to induce the farmer to exert effort or to prevent side-selling. For competition to alter

[30] For suppose that there exists a $\widetilde{q}_r \in (q/2, q)$ such that constraint (8) is satisfied. The first part of assumption 1 guarantees that the slope of the left hand side of the constraint must be steeper than the slope of the right hand side for $\widetilde{q}_r$ to exist. A contradiction. Note that the assumption can be relaxed just at the cost of keeping track of an additional case with interior solution, without gaining much further insight.

[31] The logic of the results is strengthened if farmers believe the new mill to compete through spot transactions in future seasons as well. The assumption allows a more concise exposition.

the conditions under which a relational contract is sustainable, it has to be that the side-selling constraint is harder to satisfy (see footnote below). The no side-selling constraint for the farmer is given by

$$P_r + \frac{\delta}{1-\delta}u_r \geq \frac{q\rho\mathbf{p}_0}{(\mathbf{p}_0+\rho)} + \frac{\delta}{1-\delta}u_s. \tag{9}$$

The mill will offer a relational contract in which (9) binds. Substituting for the corresponding values of u_r and u_s the binding constraint gives the minimum transfer the mill must pay to prevent competition. This is given by, $P_r \geq (1-\delta)\frac{q\rho\mathbf{p}_0}{(\mathbf{p}_0+\rho)} + \delta(\rho q/2+e)$.[32] We can, therefore, rewrite the incentive constraint of the mill as

$$\frac{\delta}{1-\delta}\left[(1+\lambda)vq_r^c - c(q_r^c) - e - \kappa - (vq/2 - c(q/2))\right] \geq P_2^c. \tag{10}$$

The mill incentive compatibility constraint under competition (10) is harder to satisfy than (8). This implies that there exists a threshold $\delta_c \in (\delta_r, 1)$ such that if $\delta < \delta_c$ either no relational contract can be satisfied at all, or the relational contract offered by the mill has a quantity q_r^c lower than optimal.

When the mill cannot offer a sustainable relational contract, the two mills compete using spot transactions. To avoid a lengthy characterization of the resulting equilibrium (which depends on the exact specification of the competition protocol), we confine ourselves to an informal discussion of the market outcomes when firms compete in spot transactions. First, since the mills are not able to offer a relational contract, there is no market price at which the farmer sells all the production at harvest. This is because the farmer has a demand for post-harvest income that spot market competition, no matter how intense, simply cannot meet. Hence, quantity sold at harvest, aggregate capacity utilization and mill efficiency must be lower than under a relational monopoly.

Second, the effects of competition on observed prices is ambiguous. On the one hand, competition between the mills implies a tendency for prices to increase. This is true both when the relational contract is sustainable (and competition simply increases the outside option of the farmer), as well as when competition destroys the relationship. On the other hand, note that when the relationship is destroyed, mills and, therefore, farmers will no longer have incentives to invest in pre-harvest inputs and effort. Since

[32] Recall the incentive constraint to induce effort is given by $P_r \geq (\rho q/2+e)$. For (9) to be the binding constraint, we must have $\mathbf{p}_0 \geq \underline{\mathbf{p}}_0 = \rho\frac{(\rho q/2+e)}{(q\rho/2-e)}$. Note, however, that $\underline{\mathbf{p}}_0$ can be smaller than $(1+\lambda)v - c(q/2)$. In other words, competition can bite on the relational contract even if the competing mill's willingness to pay is lower than the one of the mill offering the relational contract (e.g., because of transport costs).

prices under monopoly compensate farmers for the effort, observed prices at harvest might fall as a result of competition.[33] In sum,

Observation 4: *There exists a critical threshold* $\delta_c \in (\delta_r, 1)$ *such that if* $\delta \in (\delta_c, \delta_r)$ *a relational contract between the mill and the farmer is sustainable under monopoly but is not sustainable under competition. When this happens, competition destroys interlinked transactions between the mill and the farmer: 1) the mill doesn't provide inputs, 2) farmers do not exert effort to enhance quality, 3) farmers sell only a fraction of their produce to the mill (and aggregate capacity utilization decreases), 4) the mill does not pay any second payment, 5) the net present value of the price paid by the mill and the second payment might be lower than what farmers obtain in the market, 6) the mill's unit costs of processing cherries are higher than under a monopolist.*

3.6 Extensions and Discussion

The framework developed above, while highly stylized, captures the essential features of the market and illustrates the key mechanisms at work. We now discuss a number of extensions that would enrich the framework, without altering the main conclusions. First, we have abstracted from spatial heterogeneity and transport costs. The framework can easily be extended to accommodate them. When this is the case, a catchment area is endogenously defined. Under monopoly, the mill would be able to sustain relational contracts with farmers nearby, since transport costs are lower. With farmers far away, the mill might be able to only engage in spot market transactions. An increase in competition can be modeled as a reduction in the distance between two mills. This would contract the region in which the mill is able to sustain relational contracts.

Introducing spatial heterogeneity would also open the door to a richer, and more nuanced, analysis of the impact of competition on farmers' welfare. In the model above, farmers earn no rent under both spot market transactions and relational contracts (in both cases the participation constraint binds). Farmers, therefore, can only weakly benefit from competition: even when it destroys relationships with the mill, competition *might* increase prices received by farmers in a spot market. Under spatial heterogeneity, if the mill cannot (perfectly) discriminate the relational contracts offered to farmers in different locations, only the participation constraint of the *marginal* farmer binds. Inframarginal farmers earn rents from the relational contract. When

[33] We discuss below the implications for farmers' welfare.

this is the case, competition can actually reduce the welfare of (some) farmers.[34]

An extension closely related to spatial competition would focus on search costs. As is well known, in models of search there is a tendency towards inefficient entry: the marginal entrant does not take into account the externality imposed on competitors who spend resources to visit farmers that have already sold to other buyers (see, e.g., Antras and Costinot). Search costs could provide a micro-foundation for decreasing unit costs of processing/procurement. The inefficiency associated with search frictions, however, can be eliminated in models of competitive search. In these models mills post (and commit to) prices and farmers direct their search towards mills accordingly. We do not have systematic evidence on whether mills commit to posted prices or, instead, frequently renegotiate prices downwards once the farmer has brought cherries to the collection centre. In the empirical analysis we do not find a strong effect of competition on transport and procurement costs and, therefore, abstract from this channel in the model.[35]

It is well known, see e.g., Mankiw and Whinston (1986), that in industries with large fixed capital costs and business stealing effects there is a tendency towards excessive entry. While Mankiw and Whinston (1986) focus on oligopolistic competition between sellers, it is straightforward to recast their analysis in terms of oligopolistic buyers competing for inputs. Differentiation, which in our context naturally arises both from spatial competition as well as from relational contracting, mitigates the tendency towards excessive entry in the industry. In this sense, we point out that entry per se affects the degree of differentiation.[36]

Finally, we have also abstracted from mills' credit/liquidity constraints. Evidence collected through our survey suggests that at least 60% of the mills operating in 2012 were credit constrained.[37] Credit constraints exacerbate the mechanisms highlighted in the model. Under relational contracting, the mill can pay part of the costs of sourcing

[34] We believe the assumption that mills cannot perfectly price discriminate across farmers to be empirically relevant case, certainly at the level of spatial disaggregation at which we conduct the empirical analysis.

[35] We conjecture that hold-up from the mills is not a first order concern. The harvest season lasts several weeks and a mill might find it difficult to source cherries in later weeks if it holds-up farmers earlier on in the season. While temptations to hold-up increase closer to the end of the harvest, stations operating below capacity might be in a hurry to secure sufficient supply, i.e., the cost of a bargining break down also increase.

[36] Unless mills perfectly collude, in which case the aggregate quantity of cherries sourced doesn't increase with additional entry, the Mankiw and Whinston (1986) framework implies prices paid to farmers increase even when an additional entrant lowers social surplus. In the empirical analysis we fail to find support for this prediction, suggesting that additional mechanisms to those in Mankiw and Whinston are at play in our context.

[37] Using a confidential data from an international lender and an RDD design, Blouin and Macchiavello (2013) find evidence of credit constraints in a sample of over 300 mills operatin in 20 different countries.

cherries post-harvest, once sales have been realized. When relational contracting with the farmers breaks down, the mill has to pay the entire costs of sourcing cherries up-front. These funds must be borrowed, and a credit constrained mill might not be able to obtain the necessary liquidity. Credit constraints might also limit the impact of competition on spot market prices, as bidding for higher prices is more costly. Finally, competition might reduce the profits of the mill even when it doesn't destroy relational contracting with farmers. If retained earning from past profits allow firms to reduce borrowing requirements, the negative effects of competition might operate dynamically through this additional channel.

3.7 Summary of Predictions

We conclude this section summarizing the main empirical predictions.

- *The entry of an additional mill in the market might:*

    1. *Reduce the amount of interlinked transactions between the station and the farmers and, more generally, worsen the relationships between the mill and the farmers,*
    2. *Reduce capacity utilization and increase the processing unit cost of the mill,*
    3. *Fail to increase prices paid to farmers,*
    4. *The increase in processing unit costs of the mills could stem from increases in one or more of the following: i) procurement, transport costs, ii) labour costs, iii) finance costs.*

- *Moreover, as mills do not internalize the externality their entry imposes on competitors, the spatial distribution of mills will be more clustered than the one chosen by a planner and might feature excessive entry.*

4 Empirical Results

We now turn to the relationship between competition and relational contracts. First, we describe our baseline measure of competition. Second, we present OLS estimates. Given concerns that the OLS might not yield unbiased estimates, we describe the construction of our instruments for competition. This requires estimating an engineering model for the optimal placement of mills. After validating the first stage, we present the main IV results and a battery of robustness checks.

4.1 Measuring Competition

The baseline measure of competition is the number of mills within a 10 kms. radius from the mill. The reason behind the choice is as follows. Defining competition faced by a given mill requires a definition of the mill's catchment area. To fix ideas, let the catchment area be given by the area around the mill such that, if the mill bought most of the cherries produced in the area, it would operate at full capacity. A radius of 3.5 Km (measured in Euclidian distance) provides an appropriate estimate of the average mill's catchment area. Consider a conservative buffer that stretches the catchment area to a radius of 5 Km.[38] In most cases, this provides an upper bound on how far the mill purchases cherries. If all mills were symmetric, then, a mill would face competition from all surrounding mills within a 10 kms. radius. Figure 6 illustrates the distribution, across mills, of the number of mills within a 10 kms. radius. There is significant heterogeneity, with quite a few isolated mills and the average mill having 6 competitors.

In the survey, we asked the mill's manager the number of other mills that purchase inside the mill catchment area and at the nearest collection site used by the mill. The average mill faced competition from about 2 other mills at the nearest collection site and from 6 within the catchment area. The distribution plot of the survey measure is provided in Figure 6, Panel C. The correlation coefficient between the survey measure and the number of mills within 10 kms. radius is 0.77 and highly significant.

Mills are heterogeneous with respect to installed capacity, local growing conditions and availability of roads. This might suggest using a mill specific measure of competition. Competition could be defined taking into account capacity, growing conditions and local roads. Besides its simplicity, the baseline measure of competition presents an advantage from a measurement error point of view. To the extent that the baseline measure suffers from measurement error, OLS results will be biased towards zero. On the other hand, defining competition taking into account mill's specific conditions might introduce other sources of bias either upward or downward. In the interest of expositional simplicity, OLS results are presented using only the baseline measure. When presenting the main IV analysis, however, robustness checks are performed along a number of alternative assumptions and notions of competition.

[38] These figures are confirmed by survey evidence.

4.2 Competition and Relational Contracts: OLS Estimates

We begin by exploring the OLS relationship between competition and relational practices. We run regressions at the individual farmer and at the mill level. For brevity, we illustrate only the mill specification. The farmer specification also includes farmer's individual controls.[39] Denote with y_i the outcome of interest and by C_i competition experienced by mill i. The OLS specification is given by

$$y_i = \alpha + \beta C_i + \eta X_i + \gamma Z_i + \varepsilon_i, \tag{11}$$

where X_i, Z_i and Z_i^N are vectors of controls at the mill level (i) and ε_i is an error term. The vector X_i includes mill level controls: capacity installed, age and type (co-operative vs. private) of the mill. The vector Z_i includes controls for potential drivers of the mill's performance within 5 km radius from the mill. These controls include geographic (polynomials in elevation, slope, body rivers, coordinates), soil (type of terrain, suitability for coffee, density of coffee trees), social (farmer's characteristics, Gini in land around the mill, intensity of the genocide) and roads within a radius of 5 kms.[40] For sake of comparability with IV specifications, we bootstrap standard errors.

The model predicts the following practices to be part of the relational contract: *i*) pre-harvest loan and fertilizers from the mill to the farmer, *ii*) cherries sold on credit by the farmer to the mill, *iii*) (expectation of) second payments at the end of the harvest season and other forms of assistance from the mill to the farmers. These are also our main outcomes of interest. Results are presented using both farmers' and manager's answers. Table 4 presents the results. Columns 1, 2 and 3 focus on pre-harvest aspects of the relational contract. Competition negatively correlates with both farmers and manager reporting that the mill has given fertilizers and other inputs to the farmer. Columns 4, 5 and 6 focus on whether at harvest time farmers sell to the mill on credit. While all the coefficients are negative, the relationship is not statistically significant. Finally, Columns 7, 8 and 9 focus on post harvest aspects of the relational contract. We find that competition negatively correlates with farmer's expectation to receive a second payment at the end of the season and with mill reporting to help farmers in case of need. Finally, Column 10 presents results using the overall relational score that combines all practices. Taken together, the OLS results present a negative correlation

[39] Farmer's controls are age, gender, place of birth, education level, cognitive skills, distance from mill and farm size.

[40] There are 178 mills in the survey. The list of potential control variables is large and many controls are highly correlated with each other. While results do not appear to be sensitive to the exact specification, we will explore optimal specifications more formally in the future.

between competition and the use of relational contracts.[41]

The OLS results are consistent with the predictions of the model. The results, however, could be biased due to a number of concerns and cannot be interpreted as conclusive evidence of a negative impact of competition on relational contracts. Unobserved local conditions suitable for establishing relational contracts with farmers might also be (or correlate with) suitable conditions for establishing a mill. In this case the OLS coefficient is upward biased. Conversely, potential entrants might locate next to poorly run mills that score badly on relational contracts. In this case, the OLS coefficient is biased downward.

4.3 Construction of the Instrument: Entry Model

Given the concerns discussed above, we now turn to an IV strategy to investigate the causal impact of competition on relational contracts. The ideal instrument is a variable that, conditional on controls included in the model, *i*) strongly correlates with competition (first stage), and *ii*) does not influence the operations of the mill other than through its effect on competition (exclusion restriction). To construct instruments combine *i*) the spatial nature of competition embedded in the notion of catchment area defined above, and *ii*) drivers of suitability for entry (henceforth, "suitability") at the location level derived from an engineering model.[42] Conditional on suitability *within* the mill's catchment area, competition is instrumented with suitability *around* the mill's catchment area. To operationalise the idea, we need several steps:

i) Define the catchment area and the area around it,

ii) Define suitability at a highly disaggregated location level,

iii) Aggregate suitability at the catchment area and surrounding area level.

Definining Catchment and Surrounding Areas

Let us first consider the definition of the catchment area. There is a trade-off. On the one hand, we want to guarantee that conditions in the areas around the catchment area, which we use to construct the instrument, do not correlate with the mill's

[41] Similar results are obtained for pre-harvest loans and access to loans in case of help. The correlation between competition and training is insignificant. Consistently with the patterns in Figure 5, Figure 6 shows that competition positively correlated with unit costs.

[42] Duflo and Pandhe (2008) and Mobarak et al. (2012) provide early examples of papers that combine GIS data and technological features to estimate the impact of infrastructures. Our identification strategy is more closely related to ideas in the literature estimating peer effects in networks data (see, e.g., Bramoulle' et al. (2009), Acemoglu et al. (2014)) and to Bloom et al. (2014) on the effects of competition on efficiency in U.K. hospitals.

operations. This suggests defining a large catchment area. On the other hand, we want the instrument to predict the number of competing mills in the neighborood of the mill. As mills are unlikely to be affected by competition from other mills located far away, this calls for a catchment area which is not too large. Given the discussion above, we consider the 5 kms radius around the mill as "large", but "not too large", catchment area. Consequently, the instruments will be given by suitability between 5 and 10 Kms from the mill, conditional on suitability (and other controls) within 5 kms radius catchment area.[43]

Deriving Location Specific Suitability

We now derive suitability at a highly disaggregated location level. We take advantage of an engineering model for the optimal placement of mills. In the early years of the industry, a program coordinated by USAID together with engineers, agronomists and soil specialists developed an engineering model for the optimal placement of mills in Rwanda (see, USAID 2006). The model was intended to recommend suitable sites for the placement of wet mills based on a vector of characteristics to be aggregated into a score. The score summarizes at a high spatial resolution, the suitability of a given location for the establishment of a mill. The model was never implemented because the required GIS information was not available. While regulating bodies (and, possibly, investors) might have had knowledge of the criteria specified in the model, entry was not restricted to locations satisfying the criteria. Combining data available in a variety of government agencies, we have assembled all the necessary GIS information (defined at the 1 Km square) to implement the engineering model for the whole of Rwanda.

From a technical point of view, mills need to be located nearby coffee producing areas, roads and sources of water. The engineering model specified the following criteria:

Criterion #1: Outside National Parks, reserves and other prtected areas

Criterion #2: In sectors with at least 30,000 coffee trees,

Criteria #3: Within 3 km's from a spring source, at an elevation between minus 10 and minus 30 meters from the spring,

Criteria #4: Within 1 kilometer of a road.

[43]The robustness checks section explores specifications that take more/less conservative boundaries; allow for mill specific boundaries; and use alternative measures of competition.

For each 1 square km grid in Rwanda (henceforth, "grid") we define dummies for whether it satisfies each criteria or not.[44] We utilize these dummy variables (and their interactions) to predict the actual placement of mills at the grid level. There are thousands of potential grids where mills could have entered and 213 in which a mill as entered. For each grid we obtain a score summarizing the suitability of that particular location. Finally, the predicted scores are aggregated at the mill level taking averages within the mill's catchment area and area around it as defined above.

Figure 6 illustrates spatial variation in the engineering model's criteria. In Figure 6, Panel A, the dark grey grid boxes are ineligible for entry due to their land cover. The lightest green illustrates grid boxes that satisfy the number of trees necessary for entry, the brightest green areas highlight where the grid boxes satisfy all the criteria (trees, availability of water and roads). Dots depict presence of a mill.

All mills that have entered satisfy criteria #1 and #2 (see Figure 6 for the illustration for criterion #2). Grids not satisfying these two criteria are given a score equal to zero. Notice that criterion #2 is defined at the level of the sector to which the 1 Km2 grid belongs. Sectors are administrative units with an area of approximately 50 square Kms. Hence the inclusion of the corresponding dummy still allows to control for a much more highly disaggregated measure of tree density at the grid level.

Table 4 reports results for the remaining two criteria within the sample of grids satisfying criteria #1 and #2.[45] Column 1 considers presence of spring, Column 2 roads, Column 3 includes both and Column 4 the interaction between the two. In all specifications controls include polynomials in distances to spring and roads, elevation, average slope in the grid, density of coffee trees (or suitability for coffee), longitude, latitude and the interactions of these variables. Results support the engineering model. Even conditioning on a sample of eligible grids and a large number of potentially confounding factors, criteria #3 and #4 and their interaction predict the placement of mills. We use Table 4, Column 4 to predict for each grid a suitability score. Figure 6, Panel B, illustrates the predicted score from the model. Darker blue grid boxes indicate high probability of entry. Dots display the presence of a mill.

Aggregating Suitability and Discussion

Having obtained a score at each 1 square Km grid, we take averages of the score, its individual components and interactions and additional controls withing the catchment area and the surrounding area. Conditional on the score and other controls within the

[44] Criteria #3 is operationalised with a dummy requiring the grid to have at least one fifth of its area area within the relevant distance and elevation bandwidth from a spring.

[45] We also exclude built-up areas and water bodies (lakes, dams, ...).

5 kms catchment area, the instrument is given by the average suitability score within 5 to 10 kms area from the mill.

A number of important remarks are in order. First, although some of the criteria display discontinuities, due to aggregation at the mill level the identification strategy doesn't (necessarily) rely on these discontinuities. Second, although in our baseline specifications we use the suitability score as an instrument, we report robustness results in which we use individual score components (and their interactions) as instruments and show that results are virtually unchanged. The summary score presents the obvious advantage that it can easily be illustrated in Figures and Tables. Utilizing individual criteria and their interactions, however, provides us with multiple instruments and the potential to run overidentification tests. Third, since the score is a predicted variable, bootstrapped standard errors are reported.[46]

4.4 IV: First Stage, Reduced Form and Main Results

First Stage and Reduced Form

We instrument for competition using the predicted score in the area between 5-10 km. Specifically, the first stage is given by

$$C_i = \alpha + \widehat{\beta} S_i^{5/10} + \beta S_i^5 + \widehat{\gamma}_0 X_i + \widehat{\gamma} Z_i + \mu_i \tag{12}$$

where $S_i^{5/10}$ is the average predicted score in the neighbors of mill's i catchment area (5-10 km) and S_i^5 is the score inside the catchment area. The vectors X_i and Z_i includes controls as described above. The exclusion restriction is satisfied if, conditional on local costs drivers within 5 Km radius, average suitability in the 5-10 km area is uncorrelated with other determinants of mill's operations.

The predicted score strongly correlates with the number of mills within 10 km from the mill. Figure 7 illustrates the partial regression plot between the number of mills within 10 km of a mill and the instrument. Panel A is without controls and Panel controls for S_i^5, X_i and Z_i. The results show a strong first stage with a R-square of 0.744.

Figure 8 plots the reduced form relationship between the relational contracts score and the instrument. There is a strong negative and statistically significant relationship.

[46] We have tried a number of alternative specifications, including i) creating a linear score with the number of criteria satisfied in each grid box; ii) creating a linear score with weights from the engineering model documents (see USAID (2006)); iii) predicting the score using an OLS model, iv) omitting to control for tree density at the grid level. These alternative methods produce virtually identical results. By construction, our instrument only relies on cross-sectional variation across sites. Section 7 addresses concerns arising from dynamic strategic entry considerations.

between the two. The intuition behind the identification strategy is best illustrated in Figure 9. We run a reduced form regression between (the log of) operational unit costs and the average suitability scores at 1, 2, 3, ... 15 kms. from the mill. The regression includes all controls. Figure 9 reports a non parametric illustration of the estimated coefficients. A high suitability score in the proximity of the mill correlates with lower costs of operating the mills. This ought to be expected if the criteria for the engineering model are indeed relevant. Beyond a certain distance, however, a high score *positively* correlates with unit costs: better conditions are associated with worse outcomes. We argue that the score between 5 and 10 kms, rather than capturing other channels that affect mill's operations, has a (reduced form) negative impact on mill's outcomes because it attracts more competitors.

Main Results: Effects of Competition on Relational Contracts

Table 5 presents the results in a more formal regression framework. Column 1 reports the OLS coefficient and confirms the negative correlation between the relational contract score and competition found in Column 10 of Table 3.[47] Column 2 reports the first stage confirming the results in Figure 7. Column 3 reports the reduced form specification and confirms the results in Figure 8. There is a negative and highly significant relationship between the suitability score within 5 to 10 kms. and the relational contract score. Finally, Column 4 reports the IV estimates. The results confirm the negative impact of competition on relational contracts. An additional mill within 10 kms. from the mill reduces the relational contract score by 0.1 standard deviations. This effect is equivalent to a jump up or down of fifteen percentiles in the relational contracts ranking for the average mill. The comparison between Column 1 and Column 4 shows that the OLS coefficient is upward biased relative to the IV estimates. This is consistent with either measurement error, or with the source of bias in the OLS being the presence of unobserved features that correlate with both entry of competitors and with relational contracts.

Table 6 reports IV results considering the individual practices separately. Columns 1 and 2 focus on pre-harvest investments (fertilizers and inputs given to the farmer). Columns 3 and 4 focus on farmers selling cherries on credit to the mill at harvest time. Finally, Columns 5 and 6 consider post-harvest promises, in the form of second payments and help promised to the farmer in case of needs. Across all the six, we find that competition reduces the extent to which the practice is used by the mill.[48]

[47] Estimates are marginally different from those in Table 3, Column 10 since suitability score within 5 kms. is now added as control.

[48] Unreported results show that competition increases managers' reported sourcing from farmers that

The evidence suggests that competition reduces the amount of relational contracts between the mill and the farmers. Table 7 presents corroborating evidence based on questions regarding trust. The survey posed trust questions adapted from the World Value Surveys to both farmers (Columns 1 to 4) and managers (Columns 5 to 8). We asked questions about trust in co-worker, farmers, coffee collectors.[49] We find that competition lowers farmers' trust in mill's management and coffee collectors (Column 3 and 4). Similarly, we find that managers report lower trust towards farmers when there is more competition (Column 7). We also asked questions about generic trust and trust in family and neighbors. A placebo test find no effect of competition on these generic dimensions of trust.

4.5 Robustness of IV Estimates

Table 8 reports a number of robustness checks. For comparison, Column 1 reproduces the baseline specification of Column 4, Table 5. Columns 2 and 3 explore the robustness of the results to alternative sizes of the catchment and surrounding areas. Column 2 considers a catchment area with radius of 3.5 kms. and competition from mills within 7 kms. Column 3 considers a catchment area with radius of 7 kms. and competition from mills within 15 kms. The results are robust to these changes: competition has a negative impact on the relational score. The estimated coefficients are -0.136 for competitors within 7 kms. (Column 2), -0.085 for competitors within 10 kms. (Column 1), and -0.036 for competitors within 15 kms. (Column 3). The IV identifies a weighted average of the effects of additional competitors within the corresponding radius. The results imply that the impact of an additional competitor decreases with its distance from the mill. Using the observed frequency of competitors at various distances, we can parametrize the relationship between the average effect of an additional entrant and its distance from the mill. The results suggests that additional competitors beyond 10 Kms. do not affect relational contracts.[50]

Columns 1 to 3 assume exogenously given sizes for the catchment and surrounding areas. Columns 4 and 5 define the catchment areas taking into account mill's and

have received fertilizers and loans from other mills. This can be interpreted as direct evidence for the kind of poaching, or side-selling, that is the key mechanisms leading to relationship breakdown in the model.

[49] Since farmers were interviewed at the mill and there is only one manager in each mill, we did not ask farmers about their trust in the manager in order to avoid embarassement and misreporting. Following suggestions from our local enumerator team, we asked about trust in "people from Kigali". This is meant to capture attitudes towards business people with whom the farmer has a subordinate relationship.

[50] Using operational unit costs as dependent variables implies identical results.

location specific characteristics. Column 4 defines the size of the catchment area as the area around the mill that has the potential to produce twice as many tons of cherries than those needed by the mill to operate at full capacity. We match the number of trees from the census, measured at the village level, with the one square km grids around the mill. We assign to each tree an estimated production based on local geographic characteristics to reflect differences in potential production across localities. Based on the estimated potential production in the surrounding areas and on the installed capacity, each mill is assigned to one of the three groups defined in Columns 1 to 3. The results are, again, robust. Columns 1 to 4 define the mill's catchment area as the area within a certain Euclidian distance. Rwanda, however, is characterized by a hilly terrain and by highly heterogeneous density of roads across locations. Column 5, therefore, defines the catchment area to be the convex hull of the points that can be reached traveling 7 kms. by road from the mill. Results are, again, robust to this specification and very comparable to those obtained in Column 1.

Columns 1 to 5 use average suitability score in the area around the mill to instrument for competition. Column 6, instead, uses as instruments the four different criteria, averaged across all grids in the area surrounding the catchment area. The results confirm the evidence from the previous specifications. This approach provides multiple instruments to perform over-identification tests. The test fails to reject the hypothesis of exogenous instruments.[51]

Columns 1 to 6 have treat competitors symmetrically by defining competition as the number of mills within a certain area. Columns 7 and 8 explore the robustness of the results to the use of alternative measures of competition. Column 7 measures competition using the total capacity installed by mills within 10 kms. radius. While the F-test is somewhat weaker, possibly due to the noisier measure of competition, a negative effect of competition on relational contracts is confirmed. Columns 1 through 7 have measured competition exploiting differences in the number of neighboring mills within areas of a certain size. An alternative approach is to measure (the inverse of) competition by asking how far are the first N neighbors from the mill. This is done in Column 8 where competition is defined as distance to the seventh nearest station.[52] Results are robust to this different definition of competition.

[51] Multiple score instruments can be obtained, at the cost of additional assumptions, as follows. Define the catchment area to be 3.5 km radius (as in Column 2) and consider competition from mills within 10 kms. (as in Column 1). We can use the average suitability scores between 3.5 and 7 kms. and between 7 and 15 kms. as separate instruments. The results yield very similar point estimates, stronger F-test and, again, fail to reject the hypothesis of exogenous instruments.

[52] The choice of the seventh station is made for comparison with the baseline specification, since the average number of mills within a 10 kms. radius is around 6.5.

5 Additional Predictions: Mills and Farmers Outcomes

The model predicted that competition *might* decrease the use of relational contracts. The IV estimates show that competition decreases the use of relational contracts. When this happens, the model delivers a cluster of additional predictions on both mill and farmers' outcomes. This section tests those additional predictions.

5.1 Effects on Mills

Table 9 explores mill level outcomes. The model predicts *i*) higher processing unit costs and *ii*) lower capacity utilization. Columns 1 to 5 explore components of unit costs. Column 1 shows that unit costs increase by 3.5% as a result of an additional mill within 10 kms. Column 2 shows no effect on prices paid to farmers. Column 3 presents a placebo: as expected, competition has no effect on the conversion ratio from coffee cherries to processed parchment, a parameter of the production function. Columns 2 and 3 combined imply that competition has no effect on the costs of cherries. Using (1), total unit costs can be decomposed into direct cost of cherries and processing costs. Accordingly, Column 4 shows that processing unit costs increase as a result of competition by approximately 7%. Processing costs can be further decomposed into costs of labour, capital, transport, procurement and other costs. Column 5 shows that labour unit costs increase by 11%.[53] Unit labour costs increase if *i*) capacity utilization is, as predicted by the model, lower (and more irregular) and *ii*) there are costs to adjust labour to insufficient/irregular supply of cherries.

Columns 6 to 12 explore these aspects. Capacity utilization is measured as the total amount (in tons) of cherries processed during the harvest season divided by the total capacity installed. Column 6 confirms that an additional entrant reduces capacity utilization by almost 8%. This is a large effect given that the average capacity utilization in the industry is around 50%.[54] Column 7 shows that the number of weeks the mill is open is not affected by competition. Columns 8, 9 and 10, however, show that competition increases the likelihood the manager reports to have had days with too few cherries to process and with either too little or too many workers relative to the amount of cherries available.

Lower, and more irregular, capacity utilization leads to an increase in unit labour costs only if the mill cannot perfectly adjust hired labour to availability of cherries.

[53] We do not find significant effects of competition on the other sources of costs.

[54] Since we only surveyed active mills, these results do not include the extensive margin. We find that competition is also associated with the mill likelihood of operating.

Survey evidence confirms this to be the case. Firms do not completely turn down workers when there aren't enough cherries. The vast majority of seasonal workers is paid either bi-weekly or monthly, rather than daily. Stations do revise employment plans frequently (65% weekly or more often) depending on the dynamics of cherry procurement and market conditions. Hiring plans, however, are not adjusted freely. Contractual arrangements between mills and workers are somewhat sticky and also include elements of relational contracting. The employment relationship, for example, display elements of relational insurance. For example, 73% of stations answer they would turn down only some (and 12% none) of workers if there was very little cherries to process. Answers from interviewed workers confirm this. While competition affects some hiring practices (e.g., rehiring workers they already know), it doesn't affect this aspect nor the frequency of payments between mills and workers. Mills cannot perfectly adjust labour at a short notice at no cost.[55]

Competition could increase unit labour costs if mills compete for labour at harvest time. This is highly unlikely in these densely populated rural areas in which there are many more farmers available to work as employees than jobs. Columns 11 and 12 confirm that competition between stations has no impact on wage rates nor on the likelihood the manager reports to experience difficulties in hiring workers.

5.2 Effects on Farmers

The model emphasizes how competition destroys relationships by making it harder to satisfy incentive compatibility constraints. When this happens, the model predicts *i*) an ambiguous effect on prices paid to farmer, *ii*) a drop in the share of cherries sold to mills (since farmers can't rely on mill's second payments to smooth cash flows), *iii*) lower (or more expensive) input use. The simplest version of the model with homogenous farmers implies that competition always weekly increases farmers welfare. A simple extension with heterogenous farmers yield ambiguous predictions regarding the effect of competition on farmer's welfare.

Table 10 explores the effect of competition on a number of farmer's outcomes. Column 1 confirms the finding of Column 2 in Table 9: competition has no effect on prices received by farmers. Columns 2 and 5 provide direct support to the main mechanism in the model: competition reduces the share of cherries sold to mills (Column 2) and, accordingly, increases the likelihood farmers list having cash at the end of harvest as

[55] Competition does not appear to alter the relational contracts between the workers and the farmers. Unlike cherries, for which there is intense competition, there just is excessive supply of workers in the market.

one of the main advantages of home processing (Column 5). Note that the former result implies that the *aggregate* volume of cherries processed by mills might decrease as a result of additional entry. Column 3 shows that amount spent on inputs increases, possibly to compensate (unreported) reduction in the inputs received from the mill. Column 4 finds no impact of competition on yields, measured as production per tree. Finally, Column 6 aggregates a number of general questions on farmer's job satisfaction and finds a negative impact of competition on the overall score.[56] In sum, the available evidence rejects the hypothesis that farmers benefit from competition (and, if anything, shows an overall negative impact).

5.3 Effects on Quality

Finally, the model predicts that when competition leads to relationship breakdown the quality of the cherries produced might suffer. Each lot of coffee was inspected and cupped at the coffee board's laboratory in Kigali. Quality scores focus on a number of physical characteristics and defects that emerge following the roasting process. Defects in the lot can be characterized depending on their most likely origin: genetics, farmer's practices and mill processing (or a combination).

Table 11 presents the results. Columns 1 to 3 focus on physical examination. While we do not find any effect of competition on the quality dimensions which depend on the mill processing (parchment to green coffee conversion and moisture range), competition has a negative impact on bean size which can be directly linked to farmers not harvest cherries at the appropriate time (which costs effort). Among the criteria tested through cupping in the laboratory, competition has no impact on damages such as shells, floating beans and broken beans, which are under the control of the mill. However, competition increases severe insect damage which, again, originates from inadequate practices on the farm. The overall cupping score (and whether the lot gets a specialty grade) are negatively affected by competition, although the estimates are not precisely estimated.

[56] The score aggregates farmers agreement (on a scale between 1 and 4) with the following statements: *i*) "job gives me a chance to do the things I do best", *ii*) "job requires that I work very fast", *iii*) "pay is good", *iv*) in my job I learn new things, *v*) I am treated with respect, *vi*) I have a lot freedom to decide how to do my work, *vii*) I find work stressful.

6 Discussion

6.1 Dynamic Entry

The instrument for competition relies on cross-sectional variation in the suitability for mill placement around the mill's catchment area. Entry of new mills in the sector, however, happens over time. It is therefore important to pause and consider whether dynamic aspects of the entry decisions might invalidate the identification strategy. The main threat is posed by the possibility that, conditional on conditions at available locations, forward looking entrants strategically locate in areas with worse surroundings anticipating they will face lower competition in the future. This section presents three pieces of evidence suggesting that these considerations are unlikely to be of importance.

First, we can check how the order of entry correlates with the suitability score within 5 and 5 to 10 kms from the actual entry location. We expect earlier entrants to locate in places with higher score within 5 kms but to find no pattern with respect to the suitability score within 5 to 10 kms. Figure 12 lends support to this hypothesis. The figure plots the average score within 5 and 5 to 10 kms. against the order of entry. The Figure confirms that earlier entrants located in places with higher scores within 5 kms and that later entrants had to settle for locations with lower score. The corresponding trend in the average score between 5 and 10 Km is much less pronounced and likely driven by the correlation with the score within 5 kms. Indeed, once we control for the score within 5 kms, regression results show that the score between 5 and 10 Kms. does not predict the order of entry.

Second, we check whether less forward looking entrants locate mills nearby places in which there is subsequently more entry. If that was the case, the quality of a mill's management would negatively correlate with competition. Table 12 shows that competition has no effect on observable manager characteristics: age (Column 1); education (Column 2); cognitive ability, measured by simple Raven and numeracy tests (Column 3); tenure at the mill (Column 4); months worked for the station during the year (Column 5); training (Column 6); pay (Column 7) and incentives (Column 8) are all unaffected by the degree of competition. The results are confirmed when focusing on subsequent entry alone.

Finally, we exploit data from the two years following our survey and check whether, conditional on local suitability, unit costs of existing mills correlate with the location choice of new entrants. Table 13 reports the results. The unit of observation is a sector, the lowest level for which we have details about the location of new entrants.[57] The left hand side variable takes value equal 1 if a new mill has entered that sector in 2013 or 2014 and zero otherwise. We are interested to check if unit costs of existing mills in 2012 predict the location decision. Since existing mills' unit costs are observed only where mills existed in 2012, dummies for the existance of mills in the sector in 2012 are separately included. Columns 1 and 2 fail to find any correlation between the presence of mills and subsequent entry. Because Columns 1 and 2 do not control for suitability for entry at the sector level, the effect of existing capacity on additional entry is likely to be biased upward. Column 3 controls for the average suitability score in the sector and finds that *i*) the score positively predicts entry, and *ii*) existing capacity in neighboring sectors is negatively correlated with entry. More importantly, unit costs of existing mills in the sector (or in neighboring sectors) do not predict entry. Finally, Column 4 controls for the additional components of the entry model and confirms the results.

In sum, we find no evidence that mills base their strategic entry decisions on *i*) the efficiency of existing mills and/or *ii*) the suitability for further entry in places 5 to 10 kms. away from the chosen location.[58]

6.2 Quantifying Losses and Policy Implications

1. Motivate the Excercise

2. The planner's problem

3. Why it cannot be solved

4. Optimal "minimum distance" rule.

5. Policy Options: Zoning Regulations, Enforcing Minimum Distances vs. Contract Enforcement (Costa Rica)

7 Conclusion

- TO BE ADDED -

[57] Coincidentally, the average sector is approximately 60 Km2 and corresponds to an area with radius equal to 4.5 kms., very similar to the catchment area in the baseline specification.

[58] Practically, investors are unlikely to enter in certain locations taking into account, beyond local conditions, the suitability of neighbor locations in an area that extends beyond the 78 square kms. implied by our baseline measure for the size of the catchment area.

References

[1] Aghion, P., Bloom, N., Blundell, R., Griffith, R., & Howitt, P. (2005). Competition and Innovation: an Inverted-U Relationship. The Quarterly Journal of Economics, 120(2), 701-728.

[2] Andrabi, T., M. Ghatak and A. Kwhaja (2006) "Subcontractors for Tractors: Theory and Evidence on Flexible Specialization, Supplier Selection, and Contracting", *Journal of Development Economics*, April 79 (2): 273-302.

[3] Baker, G., Gibbons, R., & Murphy, K. J. (2002). Relational Contracts and the Theory of the Firm. Quarterly Journal of Economics, 39-84.

[4] Banerjee, A. and E. Duflo (2000) "Reputation Effects and the Limits of Contracting: A Study of the Indian Software Industry", *Quarterly Journal of Economics* 115: 989-1017.

[5] Banerjee, A. and K. Munshi (2004) "How efficiently is capital allocated? Evidence from the knitted garment industry in Tirupur". *Review of Economic Studies* 71(1): 19-42.

[6] Bardhan, P. (1983) "Labor-tying in a Poor Agrarian Economy: A Theoretical and Empirical Analysis", *Quarterly Journal of Economics*, 98 (3): 501-514.

[7] Bardhan, P. K. (Ed.). (1989). The economic theory of agrarian institutions (No. E10 B245). Oxford: Clarendon Press.

[8] Binswanger, H. P., & Rosenzweig, M. R. (1982). Production Relations in Agriculture. J. Develop. Stud.

[9] Bloom, N., Sadun, R., & Van Reenen, J. (2010). Recent advances in the empirics of organizational economics. Annu. Rev. Econ., 2(1), 105-137.

[10] Bloom, N., Genakos, C., Sadun, R., & Van Reenen, J. (2012). Management practices across firms and countries. The Academy of Management Perspectives, 26(1), 12-33.

[11] Bloom, N., Propper, C., Seiler, S., & Van Reenen, J. (2014). The impact of competition on management quality: evidence from public hospitals, mimeo LSE

[12] Bloom, N. and J. Van Reenen (2007) "Measuring and Explaining Management Practices across Firms and Countries", *Quarterly Journal of Economics*, 122(4), pp. 1351-1408.

[13] Bloom, N., B. Eifert, A. Mahajan, D. McKenzie and J. Roberts (2013) "Does Management Matter? Evidence from India", *Quarterly Journal of Economics*, 128(1), pp. 1-51.

[14] Blouin, A. (2014) "Culture and Contracts: The Historical Legacy of Forced Labour", *mimeo.*

[15] Blouin, A. and R. Macchiavello (2013) "Tropical Lending: International Prices, Strategic Default and Credit Constraints among Coffee Washing Stations" *mimeo*

[16] Bramoullé, Y., Djebbari, H., & Fortin, B. (2009). Identification of peer effects through social networks. Journal of econometrics, 150(1), 41-55.

[17] Casaburi, L. and T. Reed (2013) "Interlinked Transactions and Pass-Through: Experimental Evidence from Sierra Leone", mimeo Stanford

[18] Casaburi, L. and R. Macchiavello (2014) "Interlinked Transactions, Saving Constraints, and Market Structure in Kenya Dairy Farming", mimeo Stanford

[19] Diamond, P. (1987) "Consumer differences and prices in a search model", *Quarterly Journal of Economics*, 429-436.

[20] Duflo, E., and Pande, R. (2007). Dams. The Quarterly Journal of Economics,122(2), 601-646.

[21] Fafchamps, M. (2000) "Ethnicity and Credit in African Manufacturing", *Journal of Development Economics*, 61(1): 205-235.

[22] Fafchamps, M. (2004) *Market Institutions in Sub-Saharan Africa*, MIT Press, Cambridge Mass.

[23] Fafchamps, M. (2006) "Spontaneous Markets, Networks and Social Capital: Lessons from Africa", mimeo Oxford.

[24] Ghani, T. and T. Reed (2014) "Competing for Relationships: Markets and Informal Institutions in Sierra Leone," mimeo Berkeley

[25] Gibbons, R., & Henderson, R. (2012). What Do Managers Do?: Exploring Persistent Performance Differences Among Seemingly Similar Enterprises. Harvard Business School.

[26] Gibbons, R., & Roberts, J. (Eds.). (2013). The Handbook of Organizational Economics. Princeton University Press.

[27] Gil, R., & Zanarone, G. (2014). Testing for the Interaction of Formal and Informal Contracts.

[28] Ghosh, P., Mookherjee, D., & Ray, D. (2000). Credit rationing in developing countries: an overview of the theory. Readings in the theory of economic development, 383-401.

[29] Ghosh, P., & Ray, D. (1996). Cooperation in community interaction without information flows. The Review of Economic Studies, 63(3), 491-519.

[30] Greif, A. (1989). Reputation and coalitions in medieval trade: evidence on the Maghribi traders. The journal of economic history, 49(04), 857-882.

[31] Greif, A. (1993). Contract enforceability and economic institutions in early trade: The Maghribi traders' coalition. The American economic review, 525-548.

[32] Hosios, A. (1990) "On the efficiency of matching and related models of search and unemployment." *The Review of Economic Studies* 57.2: 279-298.

[33] Kranton, R. E. (1996). Reciprocal exchange: a self-sustaining system. The American Economic Review, 830-851.

[34] Kranton, R. E., & Swamy, A. V. (1999). The hazards of piecemeal reform: British civil courts and the credit market in colonial India. Journal of Development Economics, 58(1), 1-24.

[35] Kranton, R., and Swamy, A. (2008). Contracts, hold-up, and exports: textiles and opium in colonial India. The American Economic Review, 967-989.

[36] Holmes, T. J., & Schmitz Jr, J. A. (2010). Competition and Productivity: A Review of Evidence. Annu. Rev. Econ., 2(1), 619-642.

[37] Hsieh, C. and P. J. Klenow (2009) "Misallocation and Manufacturing TFP in China and India", *Quarterly Journal of Economics*, 124(4), pp. 1403-48.

[38] Hsieh, C. and B. Olken (2014) "The Missing 'Missing Middle'", *Journal of Economic Perspective*, forthcoming

[39] Hsieh, C. and E. Moretti (2003) "Can Free Entry Be Inefficient? Fixed Commissions and Social Waste in the Real Estate Industry", *Journal of Political Economy* 111-5.

[40] Ichniowski, C., & Shaw, K. (2012). Insider econometrics: a roadmap to estimating models of organizational performance. Handbook of Organizational Economics.

[41] Levin, J. (2003). Relational incentive contracts. The American Economic Review, 93(3), 835-857.

[42] Lipscomb, M., Mobarak, M. A., & Barham, T. (2013). Development effects of electrification: evidence from the topographic placement of hydropower plants in Brazil. American Economic Journal: Applied Economics, 5(2), 200-231.

[43] Little, P. D., & Watts, M. (Eds.). (1994). Living under contract: contract farming and agrarian transformation in sub-Saharan Africa. Univ of Wisconsin Press.

[44] Macchiavello, R. (2010) "Development Uncorked: Reputation Acquisition in the New Market for Chilean Wines in the UK", CEPR DP7698

[45] MacLeod, W. B. (2007). Reputations, relationships, and contract enforcement.Journal of economic literature, 595-628.

[46] MacLeod, W. B., & Malcomson, J. M. (1989). Implicit contracts, incentive compatibility, and involuntary unemployment. Econometrica: Journal of the Econometric Society, 447-480.

[47] McMillan, J. and C. Woodruff (1999) "Inter rm Relationships and Informal Credit in Vietnam", *Quarterly Journal of Economics*, 114: 1285-1320.

[48] Mankiw, N. and M. D. Whinston (1986) "Free Entry and Social Inefficiency", *The RAND Journal of Economics*, 17(1), pp. 48-58.

[49] Mukherjee, A., & Ray, D. (1995). Labor tying. Journal of Development Economics, 47(2), 207-239.

[50] Mullainathan, S. and S. Sukhtankar (2014) "Ownership Structure and Economic Outcomes: The Case of Sugar Mills in India", *mimeo*

[51] Munshi, K. (2010) "The Birth of a Business Community: Historical Disadvantage and Contemporary Mobility in India", *Review of Economic Studies*, forthcoming.

[52] Nickell, S. J. (1996). Competition and corporate performance. Journal of political economy, 724-746.

[53] Petersen, M. A., & Rajan, R. G. (1994). The benefits of lending relationships: Evidence from small business data. The journal of Finance, 49(1), 3-37.

[54] Petersen, M. A., & Rajan, R. G. (1995). The effect of credit market competition on lending relationships. The Quarterly Journal of Economics, 407-443.

[55] Porto, G. G., Chauvin, N. D., & Olarreaga, M. (2011). Supply Chains in Export Agriculture, Competition, and Poverty in Sub-Saharan Africa. Centre for Economic Policy Research.

[56] Ray, D., & Sengupta, K. (1991). Interlinkages and the Pattern of Competition.

[57] Söderbom, M. and S. Kamarudeen (2013) "Constraints and opportunities in Rwanda's industrial sector", International Growth Centre

[58] Syverson, C. (2011). What Determines Productivity?. Journal of Economic Literature, 49(2), 326-365.

[59] Syverson, C. (2004) "Product Substitutability and Productivity Dispersion", *Review of Economics and Statistics*, 86(2), pp. 534-50.

[60] Syverson, C. (2011) "What Determines Productivity?", *Journal of Economic Literature*, 49(2), pp. 326-365.

[61] World Bank (2008) "World Development Report: Agriculture for Development", World Bank

TABLE 1: DESCRIPTIVE STATISTICS, BASIC MILL CHARACTERISTICS

Variable:	Mean	Median	Private vs Cooperatives
Cooperative (=1)	0.54	1	--
Theoretical Capacity (Tons)	447	500	162.59***
End of Season Utlization Rate	47.20%	32.40%	-11.36***
Average Price Paid to Farmers (RWF)	208.53	200.00	4.44***
Output per worker	3.98	3.11	3.4**
Workers per ton of capacity installed	0.21	0.18	-0.04**
Production (Tons of Green Coffee)	46.01	32.00	18.09116***
Cherries Purchased (Tons)	294.82	199.91	120.4038***
Unit Costs [RWF per Kg]	1792.99	1800.00	97.93***
Altitude (m)	1584.16	1570.50	1.615
Historic Suitability for Coffee	0.50	0.485	0.002
Current Suitability for Coffee	0.61	0.606	-0.014
Distance to nearest paved road (km)	14.52	10.00	1.031

TABLE 2: DESCRIPTIVE STATISTICS, "RELATIONAL CONTRACTS" IN THE INDUSTRY

Variable:	Mean	SD	N
Panel A: Survey Module - Farmer			
Farmer Perspective Pre-Harvest			
Received Loans From Mill	0.173	0.378	890
Received Fertilizers From Mill	0.176	0.381	890
Has received help for unexpected expense from Mill in the p	0.250	0.433	890
Farmer Perspective at Harvest			
Sold Cherries in Advance	0.066	0.248	890
Sold Cherries on Credit	0.157	0.309	890
Farmers Perspective Post Harvest			
Expects to Receive Help	0.637	0.481	890
Expects to Receive Second Payment	0.795	0.404	890
Panel B: Survey Module - Manager			
Managers Perspective Before Harvest			
Given Loans to Farmers	0.159	0.365	178
Given Fertilizers to Farmers	0.516	0.501	178
Given Training to Farmers	0.272	0.445	178
Managers Perspective at Harvest			
Has Purchased Cherries on Credit	0.306	0.461	178
Managers Perspective Post Harvest			
Has made Second Payment in Past	0.459	0.499	178
Delay Payment (i.e. non-spot)	0.226	0.419	178
Mill provides loans to help Farmers	0.772	0.420	178

Source: All the variables are from the coffee mill and farmer survey designed and conducted by the authors in the coffee season of 2012. Panel A are *Farmer* responses and Panel B are *Manager* responses. Survey instrument and variable construction are described in the Appendix.

TABLE 3: OLS – A FIRST LOOK AT COMPETITION AND RELATIONAL CONTRACTS

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]
	PRE-HARVEST FARMER	**PRE-HARVEST MANAGER**	**PRE-HARVEST**	**HARVEST FARMER**	**HARVEST MANAGER**	**HARVEST**	**POST-HARVEST FARMER**	**POST-HARVEST MANAGER**	**POST-HARVEST**	**RELATIONAL CONTRACT**
Dependent Variable	**Received Fertilizer from Mill**	**Inputs given to Farmers**	**Z-SCORE**	**Sold Cherries on Credit**	**Has purchased cherries on credit**	**Z-SCORE**	**Expect to receive Second Payment**	**Farmer has unexpected expense does Mill provide loan**	**Z-SCORE**	**Z-SCORE**
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[9]
Competition	-0.075***	-0.048*	-0.062**	-0.003	-0.039	-0.024	-0.129***	-0.051*	-0.089***	-0.059***
	(0.024)	(0.029)	(0.024)	(0.003)	(0.0031)	(0.025)	(0.029)	(0.031)	(0.024)	(0.014)
Eng Model Criteria	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Geographical Controls	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Mill Controls	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Farmer Controls	yes	-	-	yes	-	-	yes	-	-	-
Adjusted R square	0.353	0.160	0.251	0.094	0.146	0.116	0.198	0.076	0.194	0.229
Observations	890	178	178	890	178	178	890	890	178	178

Notes: Standard errors in ( ) are bootstrapped. ***, **, * denote statistically significance at 1, 5, 10 percent, respectively. Different outcomes are reported in each column, as follows: first pre-harvest relationship outcomes are reported for the farmer (column 1) and than the mill (column 2), and finally the combined z-score for pre-harvest is reported. This pattern is next followed harvest relationship outcomes and than finally for post-harvest relationship outcomes. Lastly column 10 reports the relational z-score, which combines the pre-harvest, harvest and post-harvest z-score. The z-scores are the standardized variables, with mean zero and standard deviation of one. Competition is defined as the number of mills within a 10 kilometre radius of mill *i*. Controls in all specifications are as follows (i) Engineering model criteria (average presence of spring and road within 5 kilometre of mill), (ii) Geographical controls within 5 km radius of mill *i* (elevation (m), slope (degrees), historical soil suitability for coffee (potential producible tonnes per hectare), length of river (km), x- and y-coordinates, and tree density (iii) Mill controls are age of mill, age of mill square, type of mill (private, cooperative), capacity of mill and NGO supported (TechnoServe).

TABLE 4: VALIDATING THE ENGINEERING MODEL FOR MILL PLACEMENT

Model:	**Probit**			
Dependent Variable:	**Mill in Grid Box= 1**			
	(1)	**(2)**	**(3)**	**(4)**
Spring within 3km of grid box= 1	0.422**		0.423**	-3.931***
	(0.166)		(0.166)	(0.107)
Untarred Local Road within 1 km of grid box = 1		0.362***	0.362***	0.336**
		(0.137)	(0.137)	(0.137)
(Spring within 3km of grid box= 1)*(Untarred Local Road within 1 km of grid box = 1)				4.379***
				(0.264)
Controls: Geographical Polynomials	yes	yes	yes	yes
Controls: Geographical Polynomials & Interactions	yes	yes	yes	yes
Pseudo R2	0.120	0.121	0.124	0.125
Observations	13970	13970	13970	13970

Notes: Standard errors in ( ) are clustered at the sector level. ***, **, * denote statistically significance at 1, 5, 10 percent, respectively. The unit of analysis is 1 square kilometre (Grid Box). Sample of grid boxes with at least 30,000 coffee trees and in a grid box with potential for agriculture; grid boxes that are on land cover that is built-up, natural forest, runway or completely submerged in water body is excluded. Geographical controls are elevation, slope, historic coffee suitability, river (length), tree density in the sector, x- and y- coordinates. Cubic polynomials of the geographic controls as well as all their interactions are included.

TABLE 5: COMPETITION AND RELATIONAL CONTRACTS

	FIRST STAGE	REDUCED FORM	IV	
	[1]	[2]	[3]	[4]
Model	**OLS**	**OLS**	**OLS**	**2SLS**
Dependent Variable	**Relational Contract (Z-Score)**	**Competition**	**Relational Contract Score (Z-Score)**	**Relational Contract (Z-Score)**
Competition	-0.044**			-0.085**
	(0.015)			(0.041)
Score within 5-10 km of Mill (predict)		151.40***	-13.29**	
		(23.642)	(5.61)	
Eng Model Criteria	yes	yes	yes	yes
Geographical Controls	yes	yes	yes	yes
Mill Controls	yes	yes	yes	yes
Adjusted R square	0.170	0.762	0.159	0.210
Observations	178	178	178	178

Notes: Standard errors in ( ) are bootstrapped. ***, **, * denote statistically significance at 1, 5, 10 percent, respectively. Column 4, reports 2SLS estimation. Competition is defined as the number of mills within a 10 kilometre radius of mill i. Competition in column 4 is instrumented by the average Score (predict) from the engineering model within 5-10 km radius of mill i. All specifications conditional on controlling for the Score (predict) within 5 km radius of mill i. Controls in all specifications are (i) Engineering model (presence of spring within 3 km of grid, and roads within 1 km), Geographical controls within 5 km radius of mill i: elevation (m), slope (degrees), historical soil suitability for coffee (potential producible tonnes per hectare), length of river (km); x- and y-coordinates, length of river, roads and tree density. Mill controls are age of mill, age of mill square, type of mill (private, cooperative), capacity of mill and NGO supported (TechnoServe). Relational contract (z-score) is as defined in Table 3, column 10, by taking in account pre-harvest, harvest and post-harvest z-scores.

TABLE 6: COMPETITION AND COMPONENTS OF RELATIONAL CONTRACTS

	[1]	[2]	[3]	[4]	[5]	[6]
	PRE-HARVEST FARMER	**PRE-HARVEST MANAGER**	**HARVEST FARMER**	**HARVEST MANAGER**	**POST-HARVEST FARMER**	**POST-HARVEST MANAGER**
PANEL A						
Dependent Variable	**Received Fertilizer from Mill**	**Inputs given to Farmers**	**Sold Cherries on Credit**	**Has purchased cherries on credit**	**Expect to receive Second Payment**	**Farmer has unexpected expense does Mill provide loan**
Competition	-0.178***	-0.055*	-0.082*	-0.062*	-0.192***	-0.097**
	(0.072)	(0.033)	(0.048)	(0.036)	(0.072)	(0.066)
PANEL B						
OLS coefficient [Competition]	-0.010*	-0.025*	-0.002	-0.018	-0.038***	-0.021
	(0.006)	(0.015)	(0.006)	(0.014)	(0.008)	(0.014)
Eng Model Criteria	yes	yes	yes	yes	yes	yes
Geographical Controls	yes	yes	yes	yes	yes	yes
Mill Controls	yes	yes	yes	yes	yes	yes
Farmer Controls	yes	-	yes	--	yes	--
Adjusted R square	0.137	0.006	0.239	0.079	0.152	0.173
Observations	890 (farmer)	178 (mill)	890 (farmer)	178 (mill)	890 (farmer)	178 (mill)

Notes: Standard errors in ( ) are bootstrapped. ***, **, * denote statistically significance at 1, 5, 10 percent, respectively. Panel A, columns report 2SLS estimation. Panel B, columns report OLS estimation, with only the Competition variable being reported. Competition in Panel B is defined as the number of mills within a 10 kilometre radius of mill i. Competition in Panel A is instrumented by the average Score (predict) from the engineering model within 5-10 km radius of mill i. All specifications are conditional on controlling for the Score (predict) within 5 km radius of mill i. Controls in all specifications are (i) Engineering model criteria and Geographical controls within 5 km radius of mill i: elevation (m), slope (degrees), historical soil suitability for coffee (potential producible tonnes per hectare), length of river (km); x- and y-coordinates, presence of spring, roads, and tree density. Responses in this table are both from the Farmer and Manager's survey. Mill controls are age of mill, age of mill square, type of mill (private, cooperative), capacity of mill and NGO supported (TechnoServe). Farmer controls are age, gender, years of schooling, distance to mill, cognitive tests, membership in a cooperative and number of trees owned (log).

TABLE 7: COMPETITION AND TRUST

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
PANEL A	**Farmers**				**Manager**			
Dependent Variable	**General Trust**	**Trust in Family and Neighbors**	**Trust people from Kigali**	**Trust in Coffee Collectors**	**General Trust**	**Trust in Family and Neighbors**	**Trust farmers around Mill**	**Trust Coffee Collectors**
Competition	-0.013	-0.010	-0.086***	-0.086***	-0.071	0.045	-0.050*	-0.050
	(0.009)	(0.020)	(0.027)	(0.030)	(0.161)	(0.167)	(0.028)	(0.052)
PANEL B								
OLS coefficient [Competition]	-0.004	-0.007	-0.019	-0.028**	-0.110*	0.003	-0.020	-0.03
	(0.005)	(0.008)	(0.013)	(0.013)	(0.067)	(0.012)	(0.029)	(0.054)
Eng Model Criteria	yes	yes	yes	yes	yes	yes	yes	yes
Geographical Controls	yes	yes	yes	yes	yes	yes	yes	yes
Mill Controls	yes	yes	yes	yes	yes	yes	yes	yes
Respondent's Control	yes	yes	yes	yes	yes	yes	yes	yes
Adjusted R square	0.054	0.084	0.052	0.096	0.001	0.02	0.02	0.168
Observations	890	890	890	890	178	178	178	178

Notes: Standard errors in ( ) are bootstrapped. ***, **, * denote statistically significance at 1, 5, 10 percent, respectively. Panel A, columns report 2SLS estimation. Panel B, columns report OLS estimation, with only the Competition variable being reported. Competition in Panel B is defined as the number of mills within a 10 kilometre radius of mill i. Competition in Panel A is instrumented by the average Score (predict) from the engineering model within 5-10 km radius of mill i. All specifications are conditional on controlling for the Score (predict) within 5 km radius of mill i. Controls in all specifications are (i) Engineering model criteria and Geographical controls within 5 km radius of mill i: elevation (m), slope (degrees), historical soil suitability for coffee (potential producible tonnes per hectare), length of river (km); x- and y-coordinates, presence of spring (average within 5 km of mill i), roads (average within 5 km of mill i). Responses are from the Farmer's and Manager survey. Mill controls are age of mill, age of mill square, type of mill (private, cooperative) and NGO supported (TechnoServe). Farmer controls are age, gender, years of schooling, distance to mill, cognitive tests, member in a cooperative and number of trees owned (ln). For details on the Trust questions please see Appendix [].

TABLE 8: ROBUSTNESS CHECKS

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
PANEL A								
Dependent Variable:	**Relational Contract Score**							
Competition	-0.085**	-0.136*	-0.036*	-0.051*	-0.075**	-0.111**	-0.623*	0.025***
	(0.041)	(0.075)	(0.019)	(0.031)	(0.035)	(0.036)	(0.436)	(0.011)
PANEL B								
First Stage (Dep. Variable):	**N. of mills within 10Km**	**N. of mills within 7Km**	**N. of mills within 15 Km**	**N. of mills around catchment area**	**N. of mills within 7Km by road**	**N. of mills within 10Km**	**Capacity Installed within 10 Km (log)**	**(-) Distnace to the 7th nearest mill**
Instrument	151.40***	80.295***	320.58***	220.26***	32.638*	--	19.23***	-103.15***
	(23.21)	(18.996)	(34.761)	(20.52)	(17.309)		(2.37)	(14.35)
F-test	42.469	22.485	65.662	52.71	21.578	10.19	8.84	47.4
Instrument:	Average Score btw. 5-10 km	Average Score btw. 3-7 km	Average Score btw. 7-15 km	Av. Score around catchment area	Average score btw. 3-7 Km by road	Components of score btw. 5-10Km	Average Score btw. 5-10 km	Average Score btw. 5-10 km
Observations	178	178	178	178	178	178	178	178

Notes: Standard errors in ( ) are bootstrapped. ***, **, * denote statistically significance at 1, 5, 10 percent, respectively. Controls in all specifications are (i) Engineering model criteria and Geographical controls within 5 km radius of mill i: elevation (m), slope (degrees), historical soil suitability for coffee (potential producible tonnes per hectare), length of river (km); x- and y-coordinates, presence of spring (average within 5 km of mill i), roads (average within 5 km of mill i). Mill controls are age of mill, age of mill square, type of mill (private, cooperative) and NGO supported (TechnoServe). Column 1 reproduces the baseline specification of Column 4, Table 5. Column 2 considers a catchment area with radius of 3.5 kms. and competition from mills within 7 kms. Column 3 considers a catchment area with radius of 7 kms. and competition from mills within 15 kms. Column 4 defines the size of the catchment area as the area around the mill that has the potential to produce twice as many tons of cherries than those needed by the mill to operate at full capacity. Column 5 defines the catchment area to be the convex hull of the points that can be reached traveling 7 kms. by road from the mill. Column 6 uses as instruments the four different criteria of the engineering model, averaged across all grids in the area surrounding the catchment area. Column 7 measures competition using the total capacity installed by mills within 10 kms. radius. Column 8 measures competition as distance to the seventh nearest station.

TABLE 9: COMPETITION AND MILL'S OUTCOMES

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]
PANEL A												
Dependent Variable	**Unit Costs (Ln)**	**Cherry Prices (Ln)**	**Conversion Ratio (Ln)**	**Processing Unit Costs (Ln)**	**Unit Labour Costs**	**Capacity Utilization**	**Weeks Mill Processed**	**Days w/out enough cherries**	**Days with too many workers**	**Days with too few workers**	**Daily Wage (Ln)**	**Difficuilt in Hiring**
Competition	0.035***	-0.001	0.002	0.069**	0.144***	-7.791**	-0.229	0.056*	0.246***	0.164**	-0.012	0.0855
	(0.014)	(0.009)	(0.004)	(0.029)	(0.043)	(3.922)	(0.424)	(0.029)	(0.096)	(0.076)	(0.011)	(0.081)
PANEL B												
OLS coefficient [Competition]	0.009**	0.002	0.002	0.015	0.064***	-1.239	0.06	0.033**	0.081**	0.034	-0.004	0.028
	(0.005)	(0.004)	(0.002)	(0.012)	(0.024)	(1.372)	(0.162)	(0.016)	(0.036)	(0.028)	(0.004)	(0.030)
Eng Model Criteria	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Geographic Controls	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Mill Controls	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Adjusted R square	0.071	0.342	0.105	0.045	0.191	0.012	0.181	0.234	0.177	0.070	0.195	0.024
Observations	178	178	178	178	178	178	178	178	178	178	178	178

Notes: Standard errors in ( ) are bootstrapped. ***, **, * denote statistically significance at 1, 5, 10 percent, respectively. Panel A, columns report 2SLS estimation. Panel B, columns report OLS estimation, with only the Competition variable being reported. Competition in Panel B is defined as the number of mills within a 10 kilometre radius of mill i. Competition in Panel A is instrumented by the average Score (predict) from the engineering model within 5-10 km radius of mill i. All specifications are conditional on controlling for the Score (predict) within 5 km radius of mill i. Controls in all specifications are (i) Engineering model criteria and Geographical controls within 5 km radius of mill i: elevation (m), slope (degrees), historical soil suitability for coffee (potential producible tonnes per hectare), length of river (km); x- and y-coordinates, presence of spring, roads, and tree density. Responses in this table are from the Manager's survey. Mill controls are age of mill, age of mill square, type of mill (private, cooperative), capacity of mill and NGO supported (TechnoServe).

TABLE 10: COMPETITION AND FARMER OUTCOMES

	[1]	[2]	[3]	[4]	[5]	[6]
PANEL A						
Dependent Variable	**Prices (Ln)**	**Share Sold as Cherries**	**Amount spent on Inputs (RWF)**	**Total Yields**	**Demand for Savings**	**Job Satisfaction (Z-score)**
Competition	0.001	-0.014*	2569.060	-0.011	0.057***	-0.208***
	(0.003)	(0.007)	(1878.12)	(0.052)	(0.016)	(0.036)
PANEL B						
OLS coefficient [Competition]	0.002	-0.006	2525.01***	0.050**	0.011	-0.056***
	(0.002)	(0.004)	(727.39)	(0.024)	(0.007)	(0.015)
Eng Model Criteria	yes	yes	yes	yes	yes	yes
Geographical Controls	yes	yes	yes	yes	yes	yes
Mill Controls	yes	yes	yes	yes	yes	yes
Farmer Controls	yes	yes	yes	yes	yes	yes
Adjusted R square	0.366	0.084	0.2579	0.178	0.045	0.066
Observations	890	890	890	890	890	890

Notes: Standard errors in ( ) are bootstrapped. ***, **, * denote statistically significance at 1, 5, 10 percent, respectively. Panel A, columns report 2SLS estimation. Panel B, columns report OLS estimation, with only the Competition variable being reported. Competition in Panel B is defined as the number of mills within a 10 kilometre radius of mill i. Competition in Panel A is instrumented by the average Score (predict) from the engineering model within 5-10 km radius of mill i. All specifications are conditional on controlling for the Score (predict) within 5 km radius of mill i. Controls in all specifications are (i) Engineering model criteria and Geographical controls within 5 km radius of mill i: elevation (m), slope (degrees), historical soil suitability for coffee (potential producible tonnes per hectare), length of river (km); x- and y-coordinates, presence of spring (average within 5 km of mill i), roads (average within 5 km of mill i). Responses are from the Farmer's survey. Mill controls are age of mill, age of mill square, type of mill (private, cooperative) and NGO supported (TechnoServe). Farmer controls are age, gender, years of schooling, distance to mill, cognitive tests, member in a cooperative and number of trees owned (ln). In column 6, the components of the Job Satisfaction (z-score) are: ranking from 1-4 (4 being strongly agree) in the job I always to do what I am best at, the pay is good, I learn continuously new things, at the place I work I am treated with respect, I have a lot of freedom to decide how to do my work and I find work not stressful. Survey instrument and variable construction are described in the Appendix.

TABLE 11: COMPETITION AND QUALITY

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
Test	**Physical Examination**				**Cupping Laboratory post Roasting**				
Attribute Responsibility	**Mill and Farmer**	**Mill**	**Farmer**	**Farmer**	**Genetic**	**Mill**	**Mill**	**Mill and Farmer**	**Mill and Farmer**
PANEL A									
Dependent Variable	**Conversion Parchment to Green Coffee (%)**	**Ideal Moisture Range**	**Bean Size 16.05+ (%)**	**Severe Insect Damage**	**Shell**	**Floaters**	**Broken**	**Total Cupping Points**	**Speciality Grade (80+)**
Competition	0.015	0.003	-0.096**	0.084**	0.008	0.097	0.144	-0.24	-0.016
	(0.044)	(0.041)	(0.041)	(0.040)	(0.033)	(0.111)	(0.161)	(0.186)	(0.026)
Score within 5 km of Mill (predict)	-0.946	-8.925	13.232**	-9.345	-0.265	-18.345	-34.283	23.986	5.406
	(7.268)	(6.000)	(5.987)	(6.131)	(4.845)	(22.121)	(26.994)	(25.023)	(3.844)
PANEL B									
OLS coefficient [Competition]	-0.006	-0.019	-0.010	0.043***	0.003	0.046	0.03	0.066	0.003
	(0.018)	(0.015)	(0.015)	(0.016)	(0.012)	(0.044)	(0.093)	(0.088)	(0.012)
Eng Model Criteria	yes	yes	yes	yes	yes	yes	yes	yes	yes
Geographical Controls	yes	yes	yes	yes	yes	yes	yes	yes	yes
Mill Controls	yes	yes	yes	yes	yes	yes	yes	yes	yes
Adjusted R square	0.040	0.300	0.031	0.171	0.075	0.123	0.179	0.109	0.031
Observations	178	178	178	178	178	178	178	178	178

Notes: Standard errors in ( ) are bootstrapped. ***, **, * denote statistically significance at 1, 5, 10 percent, respectively. Panel A, columns are 2SLS estimation. Panel B, columns are OLS estimation, with only the Competition variable being reported. Competition is defined as the number of mills within a 10 kilometre radius of mill i. Competition is instrumented by the average Score (predict) from the engineering model within 5-10 km radius of mill i. All specifications are conditional on controlling for the Score (predict) within 5 km radius of mill i. Controls in all specifications are (i) Engineering model criteria and Geographical controls within 5 km radius of mill i: elevation (m), slope (degrees), historical soil suitability for coffee (potential producible tonnes per hectare), length of river (km); x- and y-coordinates, presence of spring, length of roads and tree density. Mill controls are age of mill, age of mill square, type of mill (private, cooperative) and NGO supported (TechnoServe). Representative coffee lot samples were taken from each mill in the 2012 season, these were taken to the coffee laboratory in Kigali and a team of coffee graders (4) lead by a Q grade cupper undertook a physical examination and cupped the sample for quality under the SCAA cupping protocol. See Appendix for details on the cupping protocol.

TABLE 12: COMPETITION AND MANAGER'S CHARACTERISTICS (PLACEBO)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
PANEL A								
Dependent Variable	**Age**	**Education**	**Cognitive Tests**	**Tenure at Mill (years)**	**Months work fulltime at Mill**	**Received Training?**	**Income from Mill (log)**	**Incentive from Mil**
Competition	-0.161	0.035	-0.014	0.145	-0.08	-0.039	0.040	-0.003
	(1.066)	(0.094)	(0.063)	(0.157)	(0.341)	(0.050)	(0.051)	(0.033)
PANEL B								
OLS coefficient [Competition]	-0.375	0.035	-0.008	-0.014	0.024	-0.006	-0.040	-0.002
	(0.362)	(0.094)	(0.027)	(0.063)	(0.113)	(0.015)	(0.051)	(0.011)
Eng Model Criteria	yes	yes	yes	yes	yes	yes	yes	yes
Geographical Controls	yes	yes	yes	yes	yes	yes	yes	yes
Mill Controls	yes	yes	yes	yes	yes	yes	yes	yes
Adjusted R square	0.108	0.121	0.079	0.227	0.299	0.135	0.503	0.111
Observations	178	178	178	178	178	178	178	178

Notes: Standard errors in ( ) are bootstrapped. ***, **, * denote statistically significance at 1, 5, 10 percent, respectively. Panel A, columns are 2SLS estimation. Panel B, columns are OLS estimation, with only the Competition variable being reported. Competition is defined as the number of mills within a 10 kilometre radius of mill i. Competition is instrumented by the average Score (predict) from the engineering model within 5-10 km radius of mill i. All specifications are conditional on controlling for the Score (predict) within 5 km radius of mill i. Controls in all specifications are (i) Engineering model criteria and Geographical controls within 5 km radius of mill i: elevation (m), slope (degrees), historical soil suitability for coffee (potential producible tonnes per hectare), length of river (km); x- and y-coordinates, presence of spring source, length of roads and tree density. Mill controls are age of mill, age of mill square, type of mill (private, cooperative), mill capacity and NGO supported (TechnoServe). Responses are from the Manager Survey.

TABLE 13: MILL UNIT COST AND ENTRY IN 2013 AND 2014 (PLACEBO)

	(1)	(2)	(3)	(4)
Dependent Variable:	**Entry of New Mills in the Sector for the Year 2013 and 2014**			
Unit Costs of Mill in Sector (2012)	-0.006	-0.024	0.024	0.05
	(0.172)	(0.171)	(0.168)	(0.166)
Capacity in Sector, 2012	58.96	54.62	0.067	-6.86
	(42.42)	(44.46)	(0.054)	(55.15)
Unit Costs of Mill in Neighbouring Sectors, 2012		0.08	4.589	0.08
		(0.06)	(58.68)	(0.06)
Capacity in Neighbouring Sectors, 2012		3.07	-15.74*	-16.72*
		(7.88)	(8.36)	(9.09)
Average Score in Sector			5.48***	3.43*
			(1.61)	(1.81)
Eng Model Criteria & Geographical Controls	no	no	no	yes
Adjusted R square	0.08	0.101	0.14	0.15
Observations	416	416	416	416

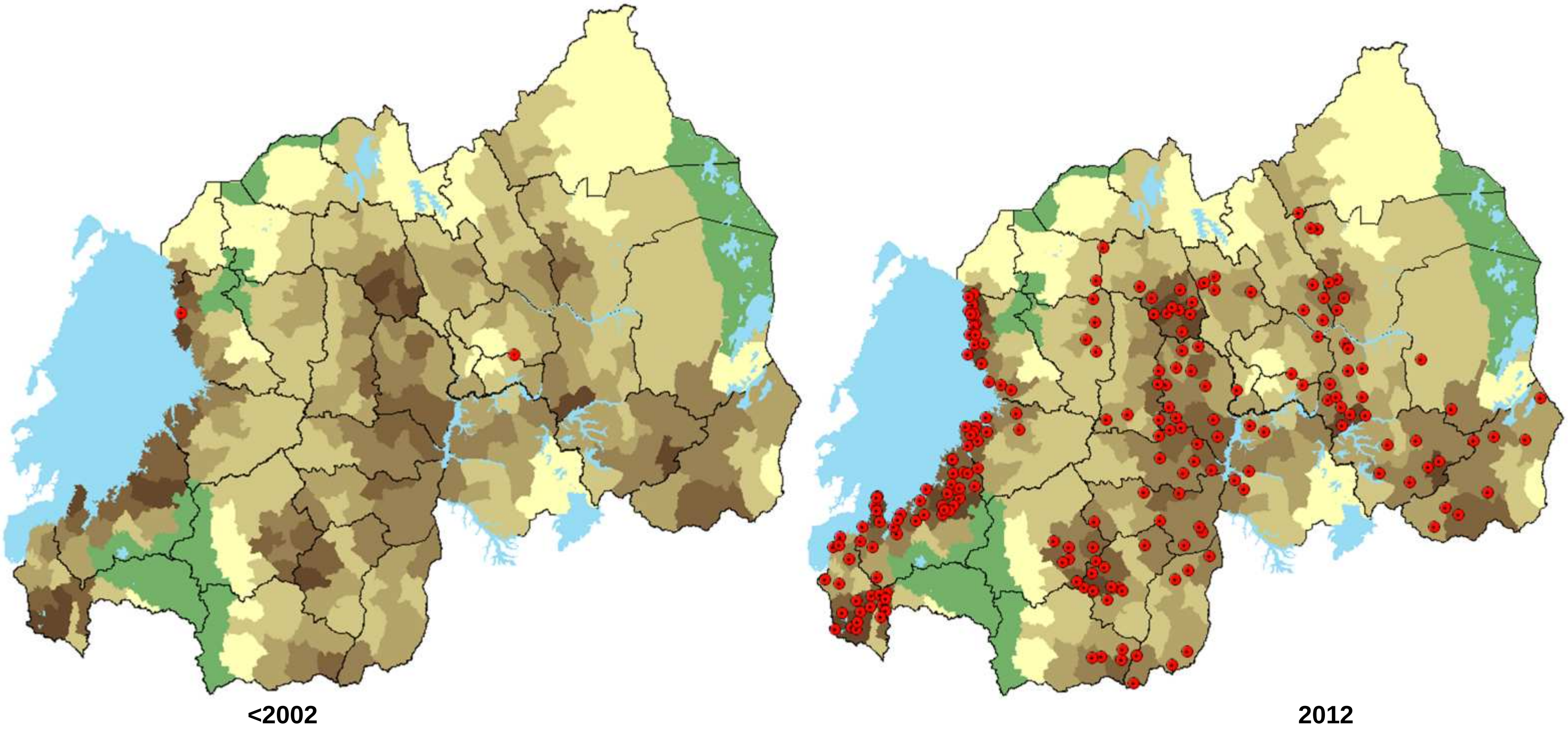
Notes: Standard errors in ( ) are bootstrapped. ***, **, * denote statistically significance at 1, 5, 10 percent, respectively. Regression unit of analysis is the Sector level (3rd layer of administration, 416 Sectors, approx. 50 sqr km). Controls in all specifications are (i) Engineering model criteria and Geographical controls within 5 km radius of mill i: elevation (m), slope (degrees), historical soil suitability for coffee (potential producible tonnes per hectare), length of river (km); x- and y-coordinates, presence of spring source, length of roads. Mill controls are age of mill, age of mill square, type of mill (private, cooperative), capacity of mill and NGO supported (TechnoServe).

TABLE A1: DISPERSION AND BREAKDOWN OF UNIT COSTS

Variable:	**Unit Costs**	**Physical Efficiency**	**Cherry Prices**	**Other Unit Costs**
p75 - p25 Ratio	1.22	1.04	1.11	1.66
p90 - p10 Ratio	1.51	1.1	1.32	2.34
% Variation explained by:				
Geography	27	34	41	30
Technology	8	6	7	4
Physical Efficiency	10	--	[4]	[3]
Cherry Prices	5	[4]	--	[7]
% Variation Unexplained:	**50**	**60**	**52**	**57**

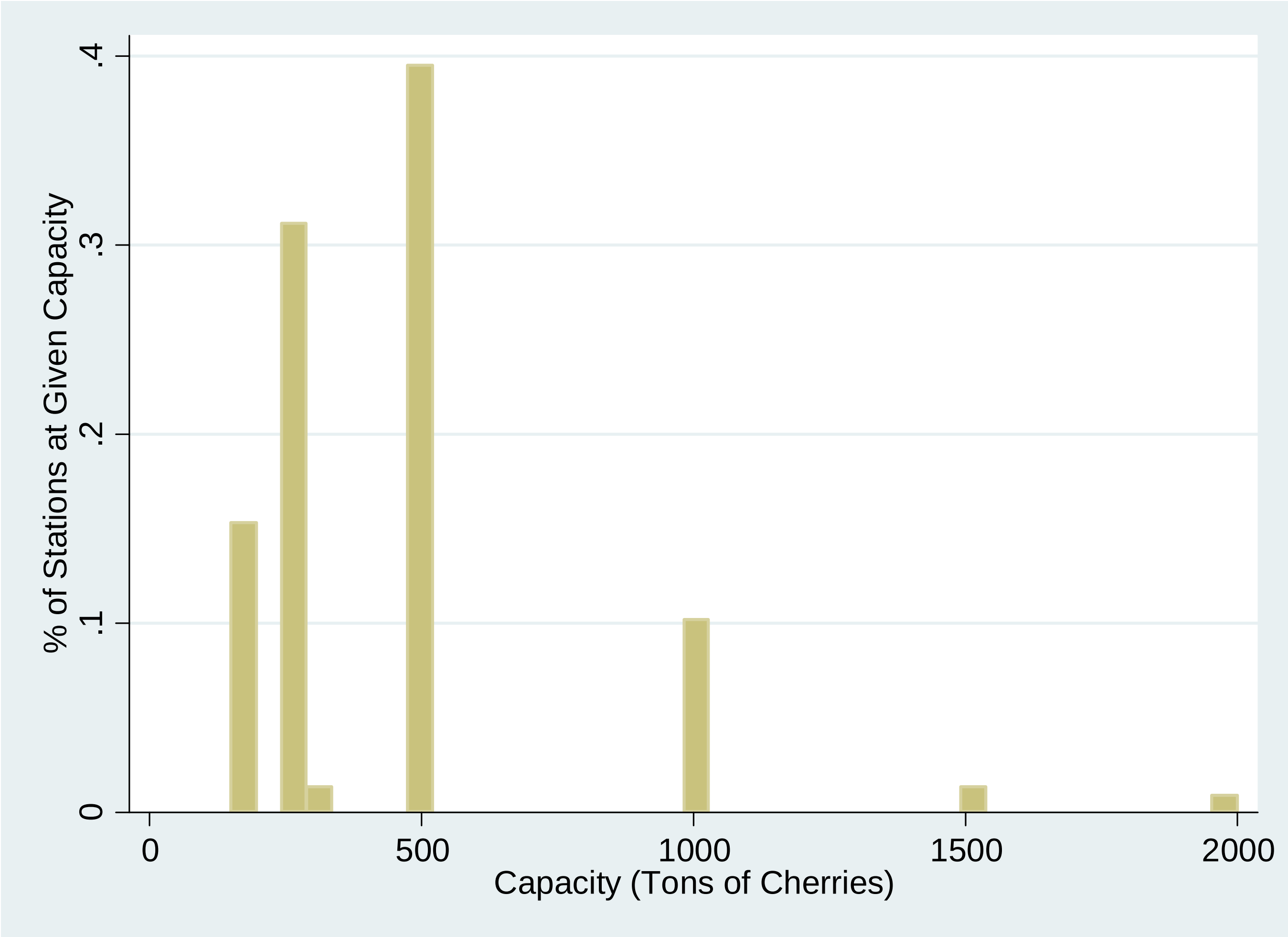
Notes: For Cherry costs: R75/25=1.15; R90/10= 1.28; above we tabulate the dispersion in terms of comparing the top 10 percentile with the bottom 10 percintle as well as the top 25 percentile with the bottom 25 percentile. The unit cost is the cost (in RfW) in processing to obtain 1 kilogram of parchment. Unit cost can be broken down into its various components - the physical efficiency (conversion of x kgs of cheeries for 1 kg of parchment, the cost of purchasing cherries and other unit costs (labor, capital, procurement and transport).

FIGURE 1: MILL PLACEMENT IN RWANDA, 2002-2012



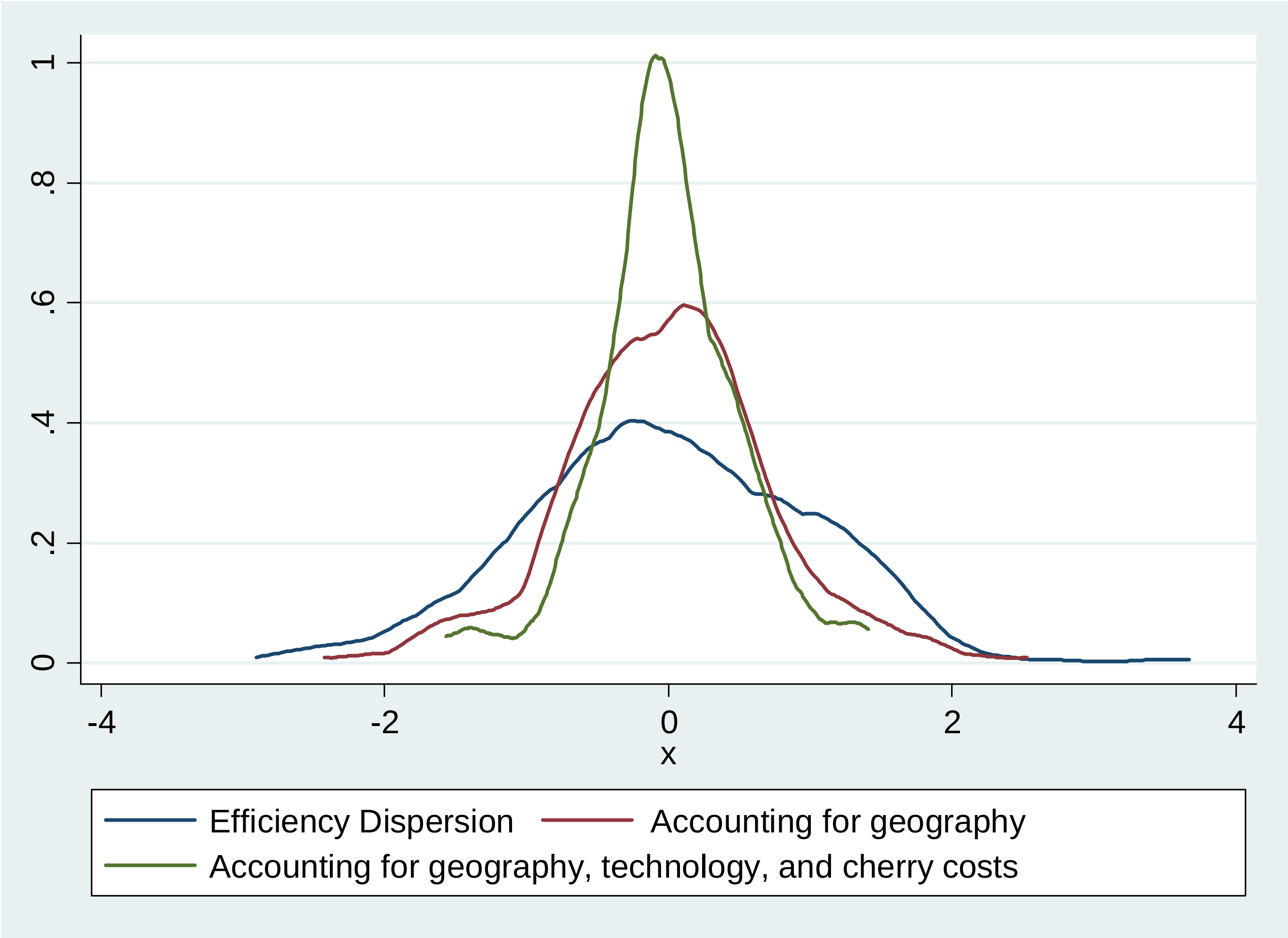
Notes: This figure illustrates, the spatial and temporal placement of mills in Rwanda. Prior to 2002 there were only 2 mills in Rwanda (denoted by red dots). By 2012 there are 214 mills built. Green shades indicate national parks, blue indicates water bodies, the background overlay is the number of coffee trees at the sector level (third administrative layer), the darker the color the higher the number of trees in the sector. Source: Authors calculation and various data sources from Coffee Board, Rwanda, see Appendix for additional details.

FIGURE 2: INSTALLED CAPACITY IN THE 2012 SEASON



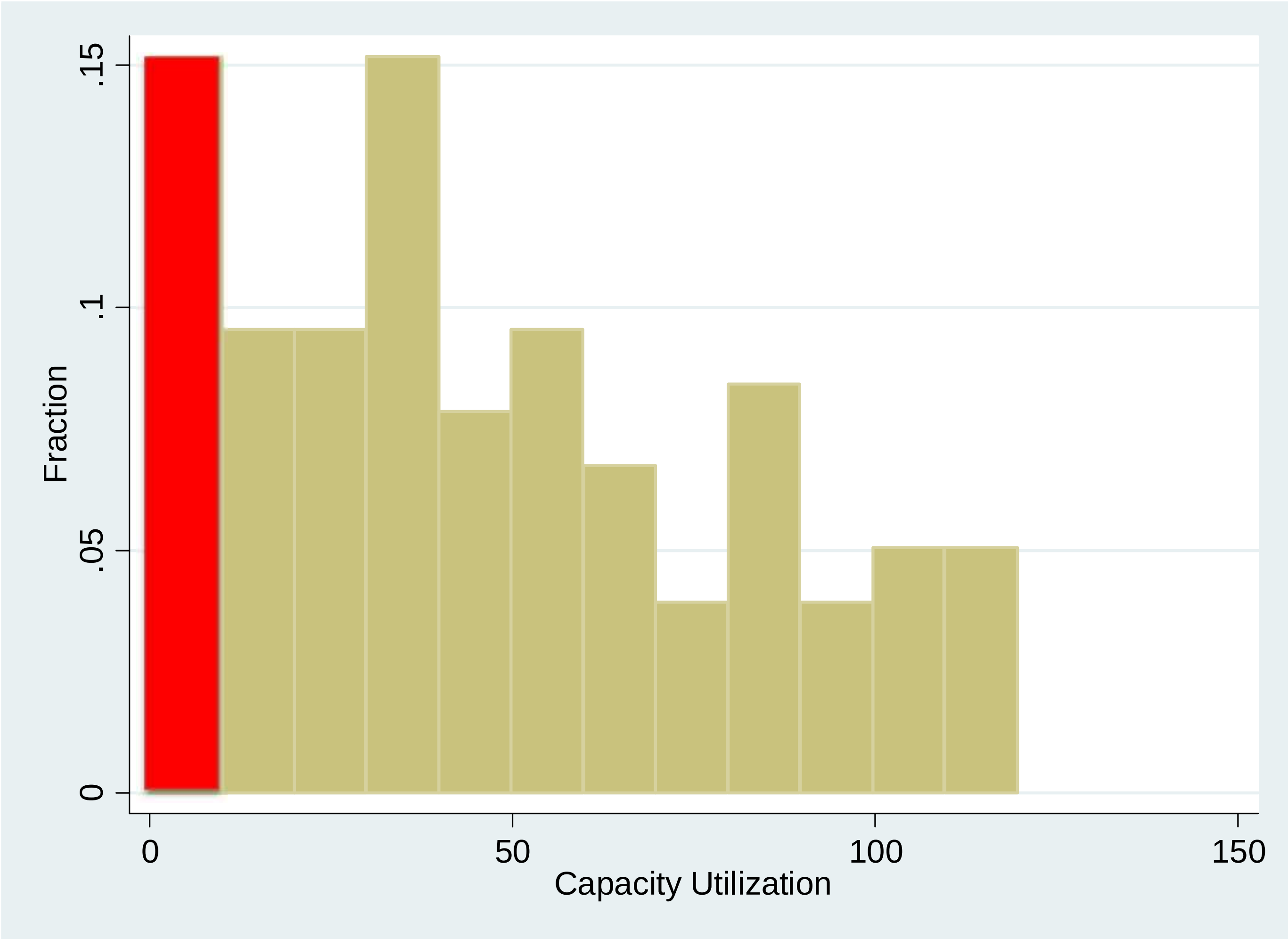
Notes: The figure plots the distribution of capacity installed by mills. The figure shows significant dispersion in capacity installed with many small stations. Source: author’s survey and Coffee Board (various), see Appendix for additional details.

FIGURE 3: DISPERSION IN TOTAL UNIT COSTS



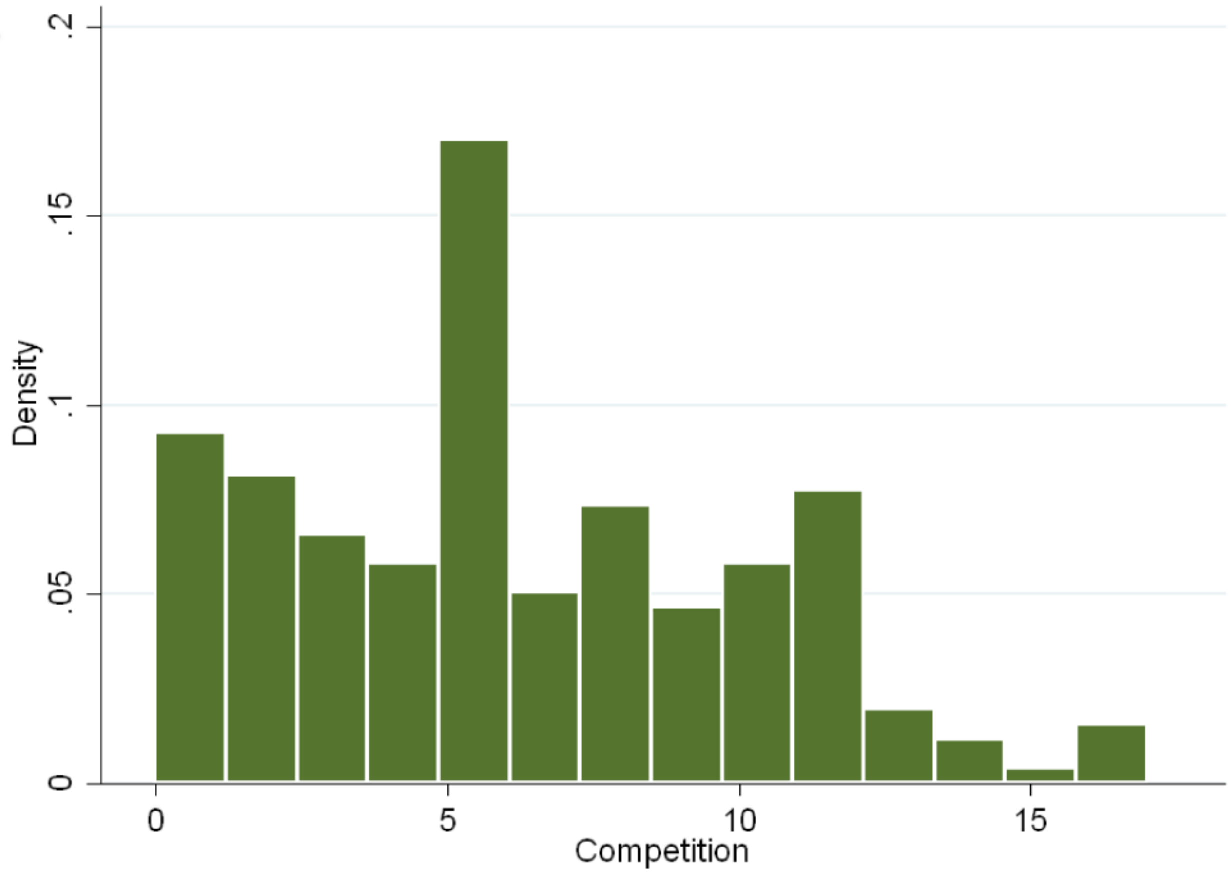
Notes: This figure shows the distribution (standardized) of efficiency, which we define as the total unit cost of obtaining one kilogram of parchment (in blue). We then control for local geography (in red) of the mill (elevation, slope, historical suitability of coffee, density of coffee trees in the sector, length of road and river and area), the type of mill (private or cooperative) and the local cost of purchasing cherries (in green). Source: Author's calculation on Mill Survey (2012) and GIS data.

FIGURE 4: DISPERSION IN CAPACITY UTILIZATION



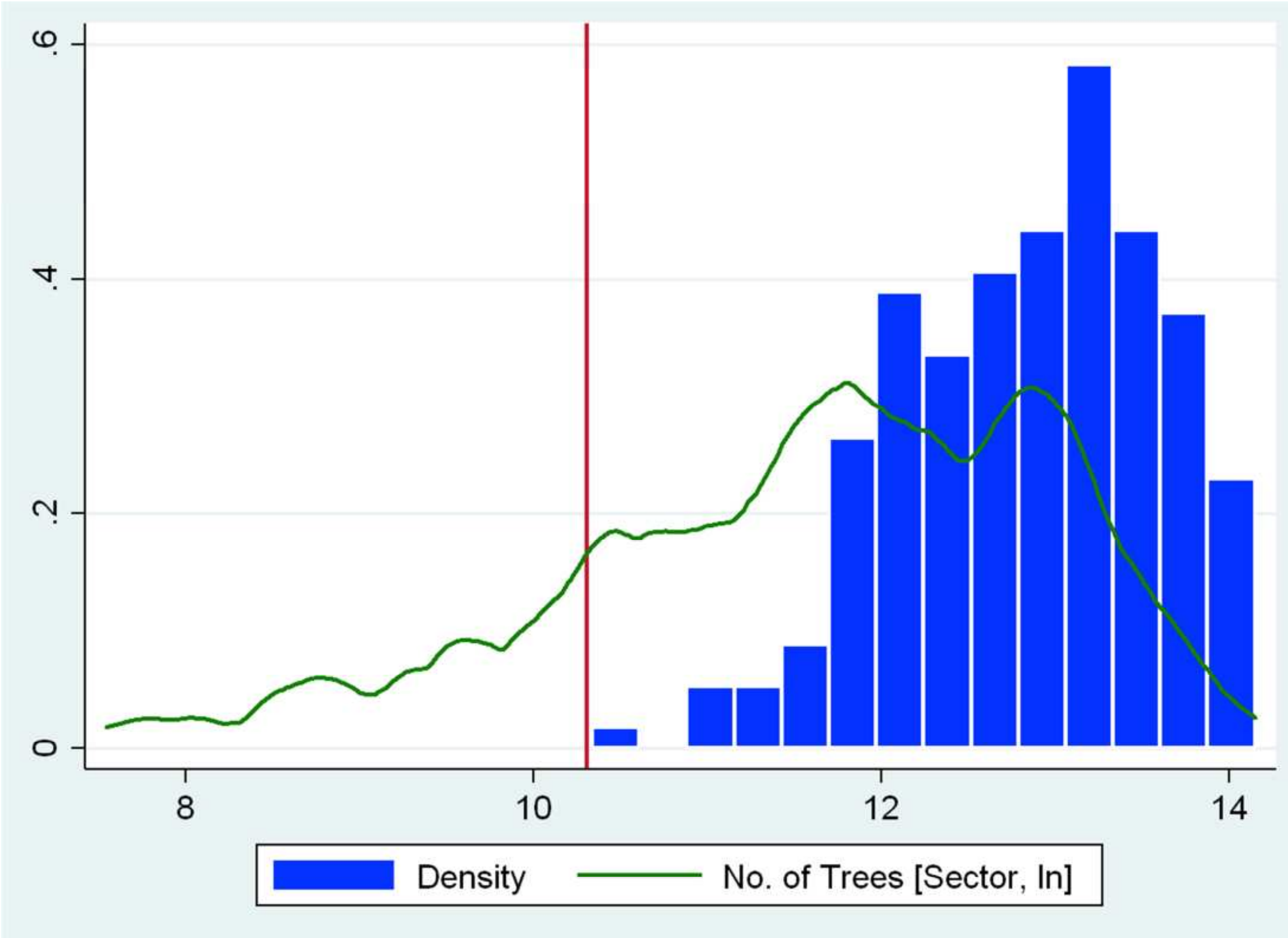
Notes: This figure displays the distribution of capacity utilization by mills for the 2012 season. In aggregate only 60% of the installed capacity is utilized. In the survey we conducted approximately 40% of the stations reported to have had days in which they wanted to purchase coffee cherries but could not.

FIGURE 6: COMPETITION BETWEEN MILLS



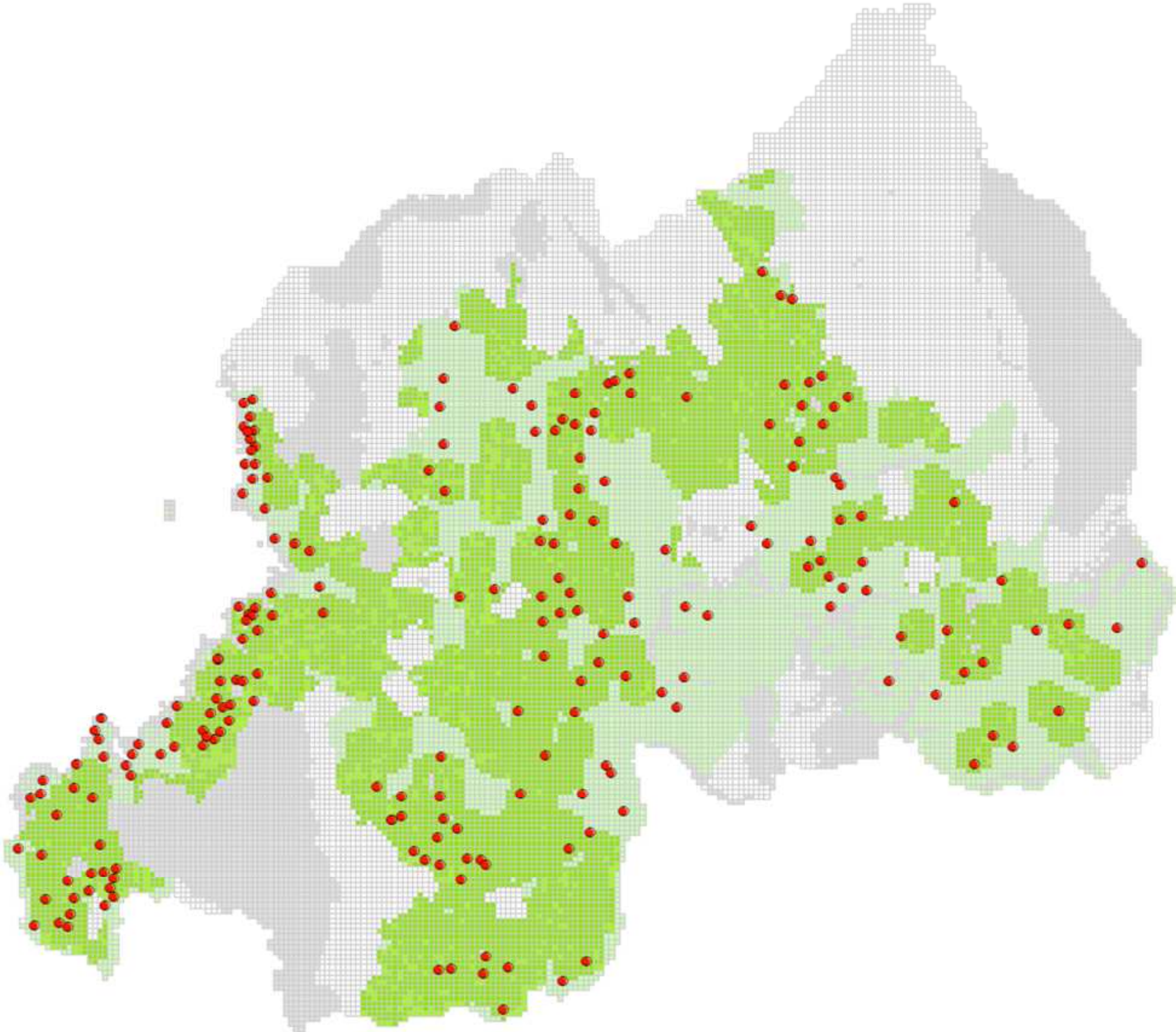
Notes: Panel A, plots the distribution of the number of mills within 10 km's of each mill, the underlying data is the geo-coded coordinates of the mills. Source: Author's calculation and Manager survey 2012.

FIGURE 7: ENGINEERING MODEL
Panel A, Mill Placement and Presence of Coffee Trees



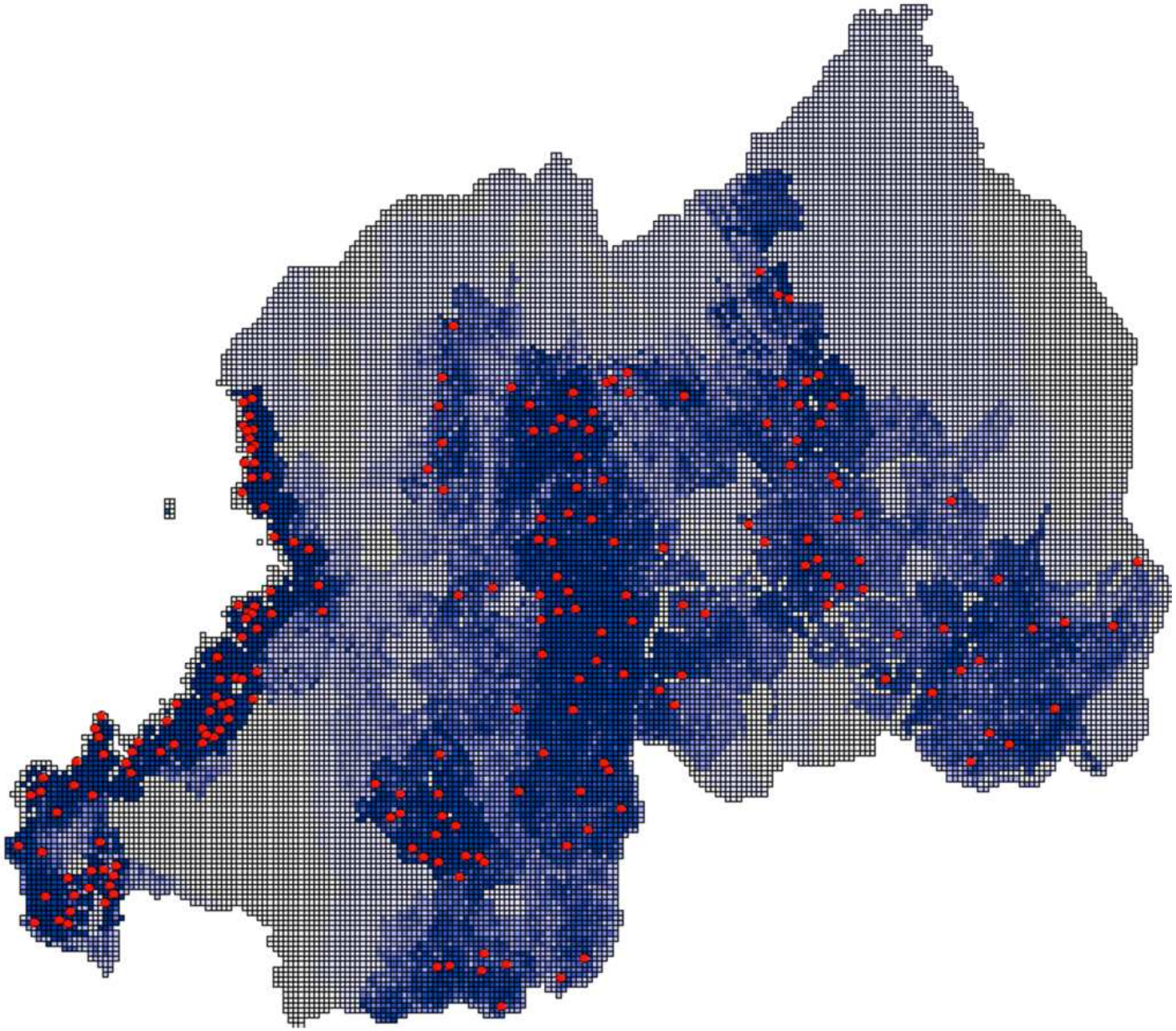
Notes: This figure plots the kernel density of coffee trees in all the administrative sectors of Rwanda (in green) that have coffee trees, and overlays on a histogram, that shows the number of coffee trees in the sector in which a mill is present (in blue). The red line indicates threshold from the engineering model, the minimum number of trees that should be available in a sector to attract a mill. *Source*: Author's calculation on Coffee Census (2009) and Mill Survey 2012.

FIGURE 7: ENGINEERING MODEL
Panel B, Criteria and Mills



Notes: This figure illustrates the engineering model's criteria; the dark grey grid boxes are ineligible for mill placement due to presence of national parks, water body or are built up areas. The lightest green illustrates grid boxes that satisfy the number of trees necessary for mill placement, the brightest green areas highlight where the grid boxes satisfy all the criteria (trees, availability of water and roads). Red dots depict presence of a mill. Source: author's calculation on various GIS datasets, see Appendix for additional details.

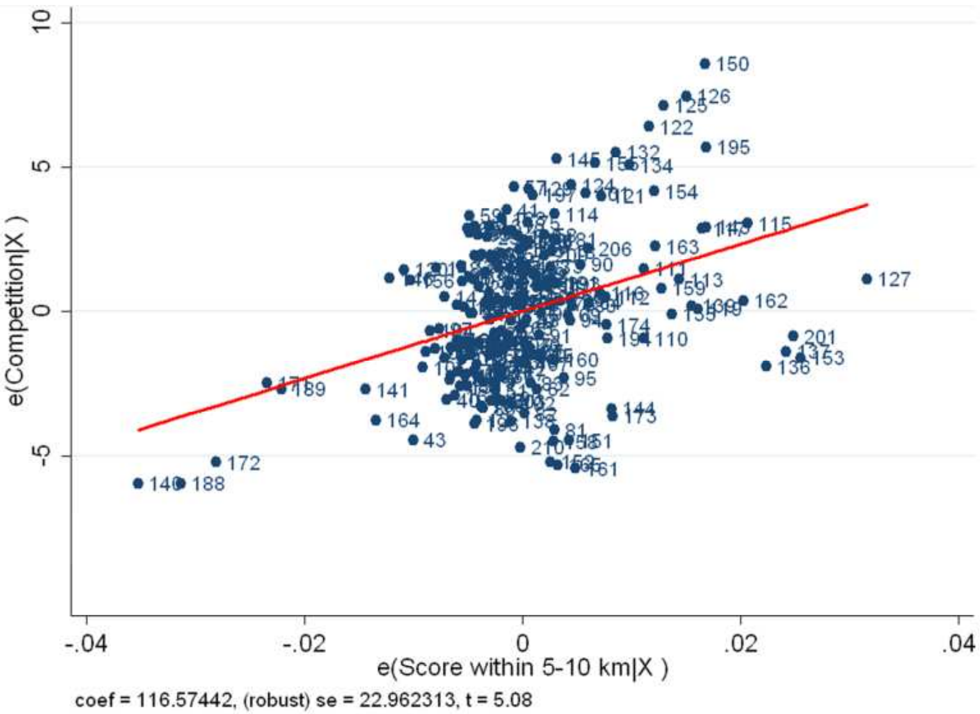
FIGURE 7: ENGINEERING MODEL
Panel C, Mill Placement



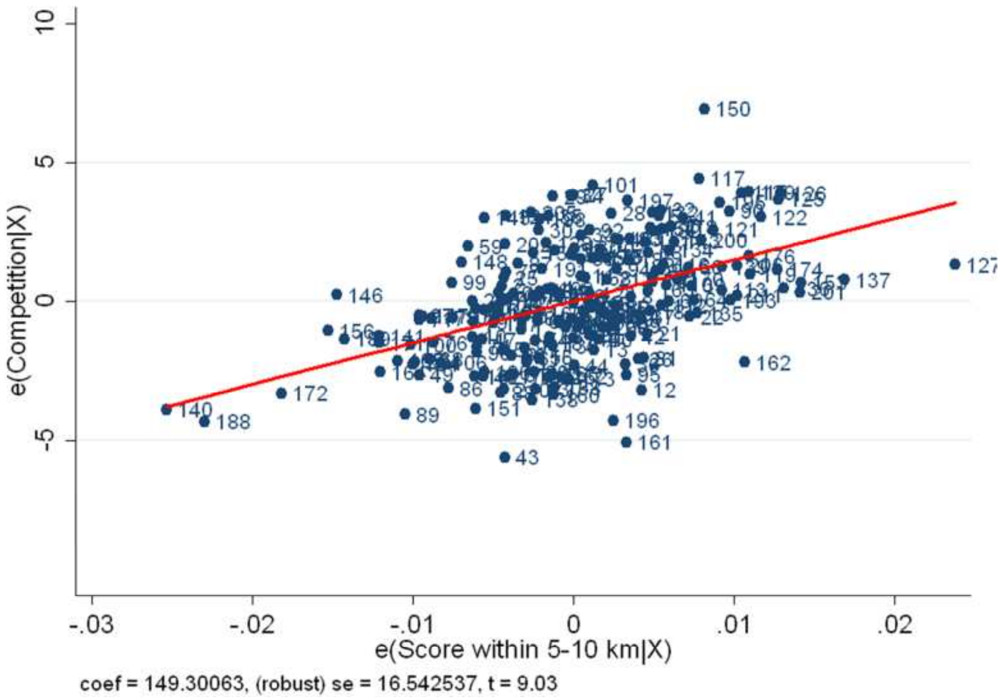
Notes: This figure illustrates, the predicted "score" for the placement of a wet mill in each grid box (1 square kilometre) in Rwanda using our model of mill placement, which is driven by engineering considerations for the optimal placement of mills. The darker the color higher the probability of mill placement. Red dots illustrate existing mills.

FIGURE 8: PARTIAL REGRESSION PLOTS, FIRST STAGE

PANEL A, WITHOUT CONTROLS

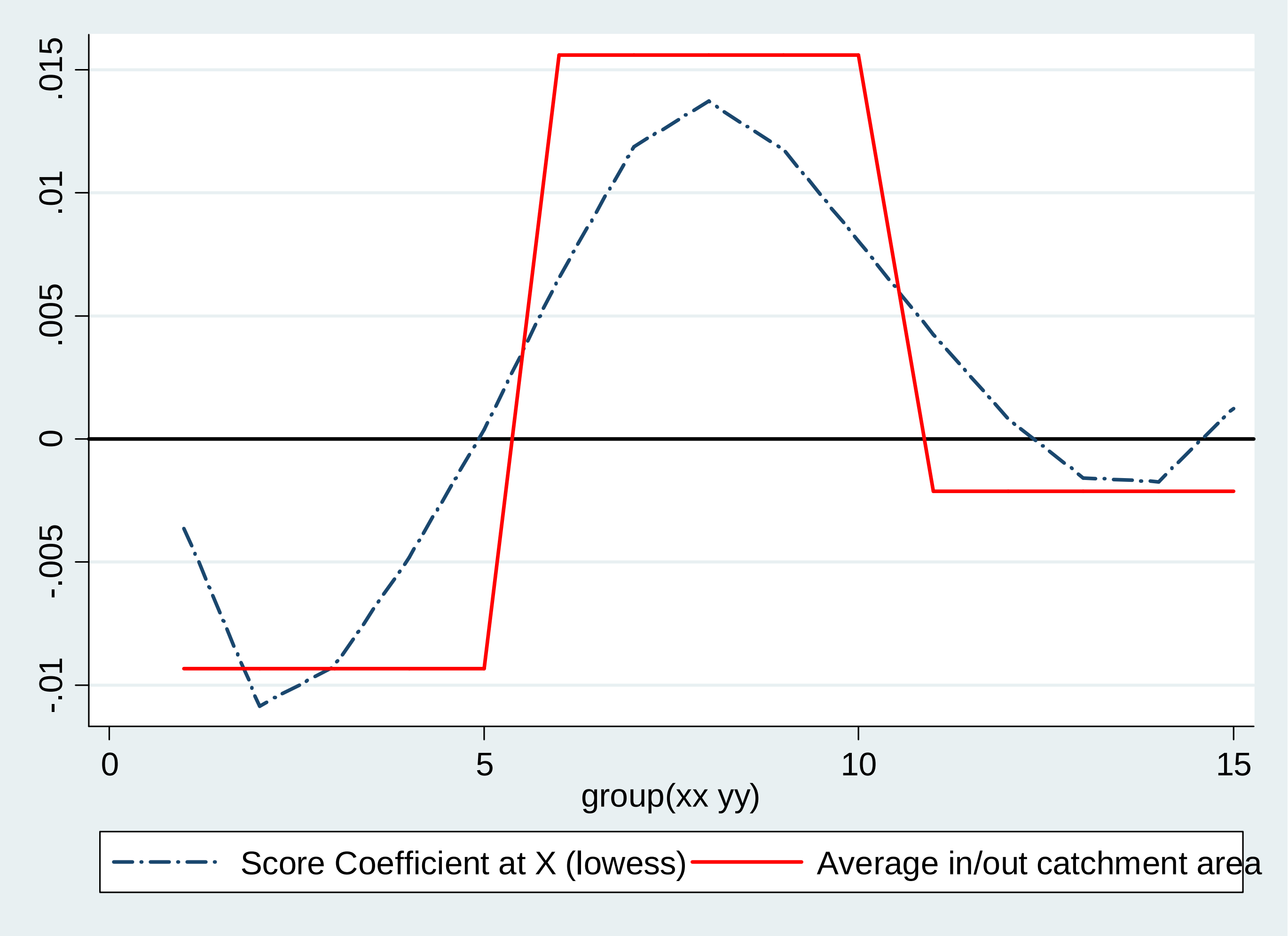


PANEL B, WITHOUT CONTROLS



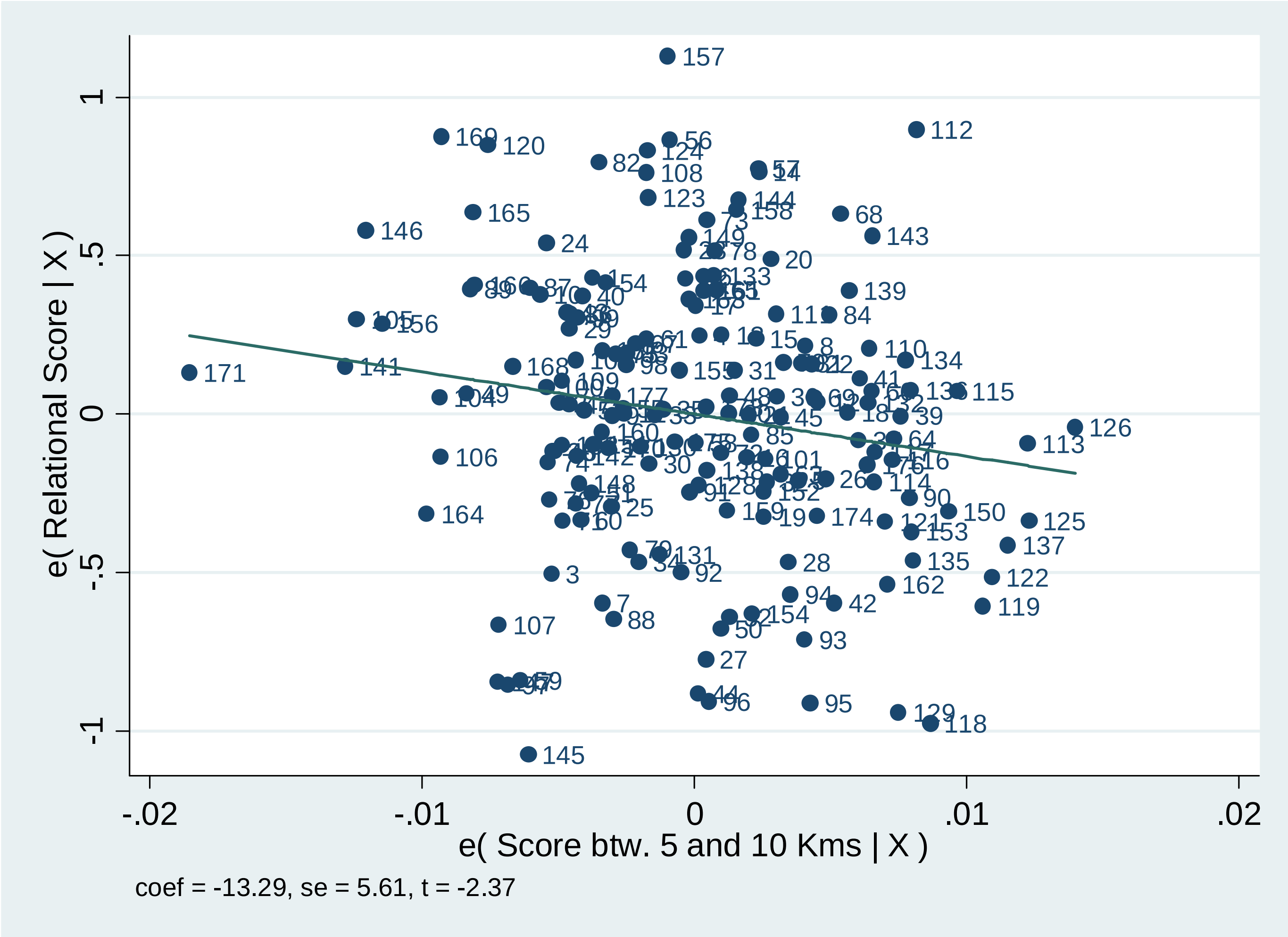
Notes: This figure plots the partial regression between competition (number of mill within 10 km of a mill) and the score within 5-10 km. Controlling for the score 5-10km and the presence of spring and roads, elevation, slope, suitability, latitude and longitude within 5 km as well as mill characteristics.

FIGURE 9: NON-PARAMETRIC ESTIMATED COEFFICIENTS



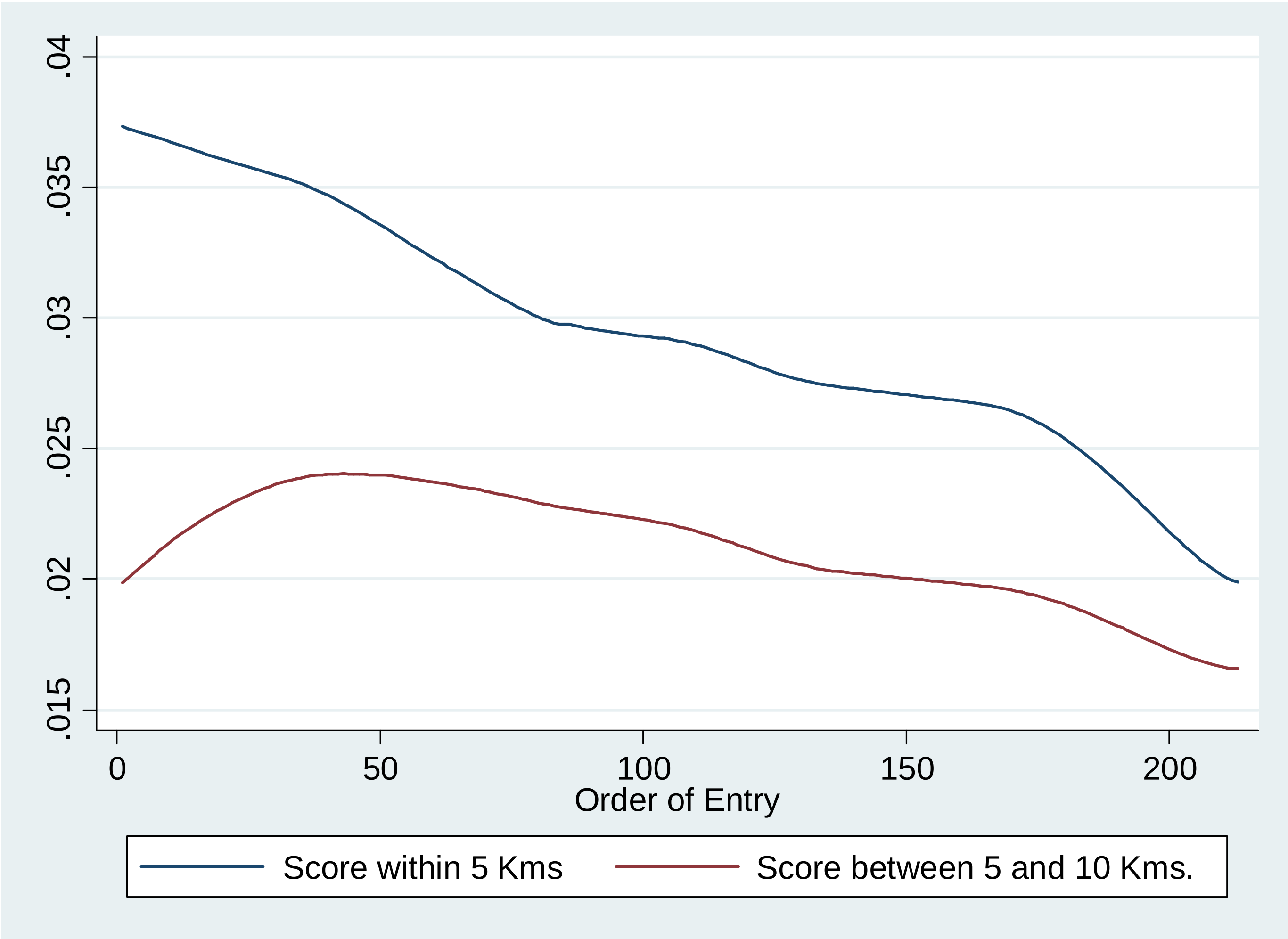
The Figure plots local polynomial estimates of a non-parametric version of the reduced form between the average suitability score at each km distance from the mill and operational costs (ln). The dotted line illustrates the estimated coefficients while the continuous line represents the average of the reduced form coefficients at 5 kms intervals, thereby reflecting the size of the catchment and surrounding areas in the baseline specification. The figure shows that a better suitability score is associated with lower unit costs in the proximity of the mill and with higher unit costs outside the catchment area.

FIGURE 10: REDUCED FORM RELATIONSHIP: BETWEEN RELATIONAL CONTRACTS AND INSTURMENT



The Figure plots the reduced form relationship between the relational score and our instrument. The Figure plots results from a partitioned regression in which all controls included in the second stage are accounted for. The Figure corresponds to Column 3 of Table 5..

FIGURE 11: ORDER OF ENTRY AND SCORE



The figure plots (lowess) average score within the catchment area (< 5 Kms) and around it (between 5 and 10 Kms) against the order of entry. The figure shows that earlier entrants located in better areas (higher average <5 Kms score) but do not appear to have chosen location according to average score between 5 and 10 Kms. Regressions results confirm that, once controlling for score within 5 Kms, score between 5 and 10 doesn't correlate with the order of entry.