# DISCUSSION PAPER SERIES

DP13607

## COMPETITION AND RELATIONAL CONTRACTS IN THE RWANDA COFFEE CHAIN

Rocco Macchiavello and Ameet Morjaria

## DEVELOPMENT ECONOMICS AND INDUSTRIAL ORGANIZATION

ISSN 0265-8003

# COMPETITION AND RELATIONAL CONTRACTS IN THE RWANDA COFFEE CHAIN

*Rocco Macchiavello and Ameet Morjaria*

Discussion Paper DP13607
Published 21 March 2019
Submitted 16 March 2019

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

# COMPETITION AND RELATIONAL CONTRACTS IN THE RWANDA COFFEE CHAIN


## Abstract

How does competition affect market outcomes when formal contracts are not enforceable, and parties' resort to relational contracts? Difficulties with measuring relational contracts and dealing with the endogeneity of competition have frustrated attempts to answer this question. We make progress by studying relational contracts between upstream farmers and downstream mills in Rwanda's coffee industry. First, we identify salient dimensions of their relational contracts (unenforceable provision of services in both directions before, during and after harvest) and measure them through an original survey of mills and farmers. Second, we take advantage of an engineering model for the optimal placement of mills to construct an instrument that isolates geographically determined variation in competition. Conditional on the suitability for mills within the catchment area, we find that mills surrounded by more suitable areas: (i) face more competition from other mills; (ii) use fewer relational contracts with farmers; and (iii) exhibit worse performance. In contrast to conventional wisdom, an additional competing mill also (iv) makes farmers worse off; (v) reduces the aggregate quantity of coffee supplied to mills by farmers; and (vi) conditional on the farmer's distance from the mill, lowers relational contracts more for farmers close to the competing mill, suggesting that competition directly alters farmers temptation to renege on the relational contract. The finding that increased competition downstream leaves all producers -- including upstream producers -- no better-off suggests a potential role for policy in a second-best environment in which contracts are hard to enforce.



JEL Classification: D43, D86, L14, O13, Q13

Keywords: Relational Contracts, Competition, Management Practices, Market imperfections, Agricultural markets



Rocco Macchiavello - r.macchiavello@lse.ac.uk
*London School of Economics and Political Sciences and CEPR*

Ameet Morjaria - a.morjaria@kellogg.northwestern.edu
*Kellogg School of Management, Northwestern University and CEPR*

# Competition and Relational Contracts in the Rwanda Coffee Chain*

Rocco Macchiavello† Ameet Morjaria‡

March 2019

**Abstract**

How does competition affect market outcomes when formal contracts are not enforceable, and parties resort to relational contracts? Difficulties with measuring relational contracts and dealing with the endogeneity of competition have frustrated attempts to answer this question. We make progress by studying relational contracts between upstream farmers and downstream mills in Rwanda's coffee industry. First, we identify salient dimensions of their relational contracts (unenforceable provision of services in both directions before, during and after harvest) and measure them through an original survey of mills and farmers. Second, we take advantage of an engineering model for the optimal placement of mills to construct an instrument that isolates geographically determined variation in competition. Conditional on the suitability for mills within the catchment area, we find that mills surrounded by more suitable areas: (i) face more competition from other mills; (ii) use fewer relational contracts with farmers; and (iii) exhibit worse performance. In contrast to conventional wisdom, an additional competing mill also (iv) makes *farmers* worse off; (v) reduces the aggregate quantity of coffee supplied to mills by farmers; and (vi) conditional on the farmer's distance from the mill, lowers relational contracts more for farmers close to the competing mill, suggesting that competition directly alters farmers *temptation* to renege on the relational contract. The finding that increased competition downstream leaves all producers – including upstream producers – no better-off suggests a potential role for policy in a second-best environment in which contracts are hard to enforce.

*This project would not have been possible without the collaboration of the Rwanda National Agricultural Exporting Board (NAEB). Without implicating them, we are especially grateful to Alex Kanyankole, Celestin Gatarayiha and Eric Rukwaya. We owe special thanks to Bob Gibbons for many helpful conversations and to Oriana Bandiera, Tim Besley, Robin Burgess, Gerard Padró i Miquel and John Sutton for their advice. We also thank Hunt Allcott, David Atkin, Nick Bloom, Gharad Bryan, Lorenzo Casaburi, Chang-Tai Hsieh, Asim Khwaja, Dilip Mookherjee, Andy Newman, Mike Powell, Steve Tadelis, and Chris Woodruff and participants at seminars in Berkeley-Haas, Bristol, BU, Boston College, Cambridge, CEMFI, Chicago Harris, Columbia, Cornell, Georgetown, Gothenburg, HBS, Harvard, Kellogg, IFS, IFPRI, Imperial, LSE, Maryland, Michigan, MIT Sloan, Notre Dame, Nottingham, Nova SBE, NYU, Oxford, SUNY Binghamton, UC Davis, UIUC, Yale SOM, Warwick, Wesleyan, World Bank and conferences at AEA 2015 (Boston), BREAD Pre-Conference (LSE), CEPR Entrepreneurship 2015, CEPR-WB GVC 2016, CSIO-IDEI NU-Toulouse 2015, CIREQ Applied Economic Conference 2017, DIME-CEGA (Kigali), EWESM 2015 (Madrid), FOM 2015 (Chicago Booth), IGC, ISNIE 2014 (Duke), NBER Org Meeting (Stanford), NEUDC (BU), SED (Warsaw), Stanford CASBS, SSDEV (Italy), 2$^{nd}$ Workshop on Relational Contracts (Madrid). Morjaria thanks NAEB for hosting numerous visits. Financial support from the IGC is gratefully acknowledged.

†London School of Economics and Political Science, *r.macchiavello@lse.ac.uk*

‡Kellogg School of Management, Northwestern University, *a.morjaria@kellogg.northwestern.edu*

# 1 Introduction

Markets in developing economies are often portrayed as dysfunctional: thin, scarcely competitive, and harboring unproductive firms. This suggests an important role for increased competition in improving firm performance and management via both selection and incentives (Syverson 2004, Bloom and Van Reenen (2010), Bloom et al. (2015)). Yet, these same markets are also often characterized by weak contract enforcement (Greif (1993), Djankov et al. (2003)). This generates an important role for *relational contracts* − informal agreements sustained by the future value of the relationship (Baker et al. (2002)). In settings with limited competition, but also weak contract enforcement, the effects of increased competition on firm performance are then theoretically ambiguous: on the one hand, competition might improve a firm's performance; on the other hand, by reducing profits and tempting parties with alternative trading opportunities, it may weaken relational contracting and reduce efficiency (Kranton (1996) and Ghosh and Ray (1996)). What is the impact of competition in such second-best institutional environments?

Answering this question has been challenging for two reasons: first, relational contracts are implicit and context-specifc and thus difficult to *measure*; second, *identification* of the causal effects of competition is complicated by the endogeneity of market structure. This paper identifies the effect of increased competition on firm outcomes in a weakly institutionalized environment in which relational contracts are needed to sustain trade. We resolve the two challenges by studying relational contracts between upstream farmers and downstream mills in Rwanda's coffee industry, a context that affords us progress in both measurement and identification.[1]

The context allows us, first, to identify specific, salient dimensions of relational contracts. Mills operate a simple technology but are affected by contractual difficulties typical of developing countries' agriculture (see, e.g., Bardhan (1989)). Due to lack of well-functioning input and financial markets, mills and farmers bundle the sale of coffee cherries at harvest with legally unenforceable provision of services in both directions before, during and after harvest. We measure the use of relational contracts between mills and farmers by conducting an original survey of all mills in the sector.

Second, we construct an instrument for competition, taking advantage of an engineering model that specifies detailed criteria for the optimal placement of mills. The instrument isolates geographically determined variation in the presence of mills which

[1]Coffee is the main source of income for about 25 million farmers worldwide and features many aspects common to other agricultural chains in developing countries.

we argue affects relational contracting only through the intensity of mill competition.

We find that, conditional on the suitability for mills within the catchment area, mills surrounded by more suitable areas: (i) face more competition from other mills (first stage); (ii) use fewer relational contracts with farmers (reduced form); and (iii) exhibit worse performance. The negative impact of competition on mill's performance could be due to standard scale effects. However, in contrast with the standard argument, we also show that an additional competing mill (iv) makes *farmers* worse off; and (v) reduces the *aggregate* volume of coffee supplied by farmers to mills. Furthermore, (vi) conditional on the farmer's distance from the mill, an additional competitor lowers the use of relational contracts more when the farmer is closer to the competing mill. These findings suggest that competition hampers relational contracts in two ways: by the usual logic, competition lowers mill's profits and thus the ability to sustain relationships; but it also directly alters farmers' *temptation* to renege on the relational contracts.

These findings must be interpreted cautiously. We identify the effect of an additional competitor for a mill that competes with five other mills on average. Our findings are thus not in conflict with Adam Smith's remark that "monopoly is a great enemy to good management". The finding that increased competition downstream leaves all producers – including upstream producers – no better off, however, provides novel evidence on the functioning of markets in second-best environments (Rodrik (2008)). In particular, it suggests the possibility of socially excessive entry when contracts are hard to enforce and a potential role for policy to improve market functioning.[2]

The paper proceeds as follows, section 2 provides industry background and presents our index of relational contracts between mills and farmers. Due to the lack of well-functioning input and financial markets, mills and farmers bundle the sale of coffee cherries with other services exchanged before, during and after harvest. We focus on three relational practices: inputs and loans provided by the mill to the farmers *before* harvest; coffee sold on credit by farmers to the mill *during* harvest; and assistance from the mill to the farmers unrelated to (i.e, *post*) harvest. Farmers and mills, however, do not have access to legal enforcement and must resort to relational contracts. In our context a relational contract is, thus, a non-legally binding agreement between a mill

[2]Business stealing can induce excessive entry even when contracts are enforceable (Mankiw and Whinston (1986)). In Mankiw and Whinston (1986) however an additional entrant increases the aggregate volume of coffee supplied by farmers to mills and makes farmers weakly better off. In contrast, our results imply an additional externality through which an entrant directly lowers other mills efficiency by destroying relational contracts with farmers. This difference explains the worse outcomes for farmers and the lower aggregate amounts of coffee processed.

and supplying farmers that describes how farmers and mills should behave over the course of the entire season.

We measure the use of each relational practice surveying managers and random farmers at each mill. We aggregate the mill's manager and farmers' responses into a relational index. The relational index correlates well with mills' capacity utilization and processing unit costs giving us confidence that it measures the relevant practices for this industry.

Section 3 presents a theoretical framework that captures the key aspects of the relationship between mills and farmers. Competition *might* reduce parties' ability to sustain relational contracts. Competition operates through two distinct channels. First, it reduces mills profits and indirectly makes it harder to sustain relationships with farmers. Second, it also directly increases farmers' temptation to renege on the relational contract. When competition leads to the breakdown of relational contracts, the model predicts: (1) a decrease in the use of *all* relational practices; (2) a decrease in mill's efficiency, capacity utilization and output quality; (3) a decrease in coffee sold for processing by, and access to inputs for, farmers; and (4) an ambigous effect on prices received by farmers and a negative effect on welfare for at least some farmers.

The empirical analysis proceeds in three steps. Section 4 asks *whether* competition breaks relational contracts; Section 5 explores the *consequences* of relational contracts breakdown; and Section 6 investigates *mechanisms.*

We measure competition as the number of mills within a 10 km radius from the mill.[3] Competition negatively correlates with the relational index. The OLS estimates of the causal impact of competition on the relational index are however likely biased. In particular, unobservable factors might correlate with competition and with the desirability, or the feasibility, of relational contracts; competitors might locate near mills with either worse or better relational practices; and competition itself is measured with error.

To address these concerns, we implement an instrumental variable strategy. We need a variable that, conditional on controls, correlates with competition (first stage) and only influences mill's operations through its effect on competition (exclusion restriction). We construct our instrument combining the spatial nature of competition with an engineering model for the optimal placement of mills in Rwanda. In the early 2000s, when only a handful of mills were established, a team involving engineers, agronomists and soil specialists developed an engineering model to identify suitable

[3] Coffee cherries must be processed within hours of harvest and roads are often in poor conditions. Mills thus mainly compete with nearby mills. Survey evidence validates our measure of competition.

sites for mills construction. The model, however, was never implemented because the required GIS data were not available at the time. Subsequent entry of mills was thus not restricted nor limited to locations satisfying the model's criteria. We assembled ex novo the data required for the model and computed a suitability score for a mill's placement at the $1km^2$ resolution for the whole of Rwanda. We aggregate the score in the region within a 5 km radius (catchment area) and in the surrounding area between 5 and 10 km from the mill (instrument). The exclusion restriction is satisfied if, conditional on suitability within the mill's catchment area, suitability in the surrounding area affects mill's operation only through its impact on competition.[4]

The IV results show a negative impact of competition on relational contracts. An additional mill within 10 km distance reduces the relational index at the mill by 0.25 standard deviations. The impact of an additional mill decreases with its distance from the mill and vanishes at approximately 10 km. Non-parametric estimates suggest that the impact of an additional competitor is small when there are few competitors and becomes stronger as competition intensifies.[5]

Section 4 concludes with showing competition leads to a breakdown in relational contracts between mills and farmers. When this happens the model yields additional predictions that are tested in Section 5. The first prediction is that, although competition only provides farmers the ability to sell coffee to a different mill *during* harvest, the breakdown in the relational contract makes *all* relational practices unsustainable. The evidence supports this prediction: an additional competing mill reduces relational practices *before*, *during* and *after* harvest by a nearly identical magnitude.

Competition also leads to worse mill-level outcomes. Mills suffer lower and more irregular capacity utilization. This leads to a 7% increase in unit processing cost, largely driven by higher labour costs. By reducing farmers' access to inputs and reducing incentives to exert effort, the breakdown in relational contracts also leads to lower output quality. We lab-test random samples of coffee produced by each mill and detect a negative impact of competition on coffee quality, particularly on those dimensions that depend on farmers' growing and harvesting practices.

The negative impact of competition on mill's outcomes is perhaps not surprising and could, in principle, reflect standard scale effects associated with business stealing

[4] The instrument yields a strong first stage. Conditional on suitability within the mill's catchment area, a one standard-deviation increase in the instrument is associated with about 2 additional competing mills. Section 6.2 discusses threats to, and evidence in support of, the identification strategy.

[5] We explore the robustness of the IV findings to alternative definitions of catchment areas and competition; alternative strategies to construct the instrument; and alternative assumptions on the structure of the error term.

(Mankiw and Whinston (1986)). In contrast to this logic, however, we also show that competition *lowers* the amount of coffee that farmers sell to any mill. In fact, competition is associated with no benefits, and possibly harmful effects, for farmers. Competition is associated with negligible increases in prices received by farmers from mills and no increase in either yields or investments in coffee plantations. Competition thus does not increase farmers' returns from coffee cultivation. Farmers lack access to adequate saving technology and often process parts of their coffee at home to sell it when they need cash. When relational contracts break down we find that farmers no longer trust mills to make payments after harvest. Accordingly, competition increases the share of coffee processed at home and the likelihood of farmers reporting savings as the main reason for home processing. At the same time, competition does not increase aggregate input use but increases the likelihood that farmers have to pay for the input themselves. Competition lowers an overall index of farmers' satisfaction.[6]

Section 6 investigates two mechanisms through which competition leads to the breakdown of relational contracts. We distinguish the effect of competition from mills that are near the mill *and* the farmer (direct competition) from the effect of competition from mills that are near the mill, but not the farmer (indirect competition). We find that indirect competition lowers the use of relational contracts with the farmer. This is consistent with competition reducing mill's profits and making it harder to sustain relational contracts with farmers. We also find, however, a stronger effect associated with direct competition. This provides suggestive evidence that competition also undermines relational contracts through a direct *temptation* mechanism that bites precisely when contracts are hard to enforce. Section 7 offers concluding remarks and discusses policy implications.

The paper contributes to three literatures. First, we contribute to the literature on relational contracts and, more broadly, on management practices (Bloom and Van Reenen (2007, 2010)).[7] The work by Bloom and Van Reenen (2007, 2010) shows that the adoption of certain well-codified management practices is strongly associated with firm's performance. This evidence raises the question of why many firms fail to adopt these management practices. A possibility is that a firm's ability to introduce, and benefit from, these management practices depends on relational contracts in place both within and across the firm's boundaries (Baker et al. (2002), Gibbons and Henderson (2012), Helper and Henderson (2014)). Systematic evidence on the prevalence and

[6]Unreported results show that competition reduces trust between mills and farmers as elicited through standard World Value Survey questions.

[7]Gil and Zanarone (2015) provide an excellent review of empirical work on relational contracts.

drivers of relational contracts can expand our understanding of management but remains scarce. This is because relational practices are, by definition, hard to codify, context-specific and, therefore, hard to measure. The paper offers an illustration of how relational practices can be systematically measured in a large sample of firms. We also show significant dispersion in the adoption of relational practices; that relational practices are complementary; and confirm that their adoption correlates with firms' performance.

Second, we study the effect of competition on market functioning in an environment characterized by poor contract enforcement.[8] There is abundant evidence that competition is associated with higher productivity through both selection and incentive effects (see Holmes and Schmitz (2010) for a survey). For example, Syverson (2004) shows that in the US larger, more competitive markets, are associated with stronger selection in concrete manufacturing. Schmitz (2005) shows that, in response to competition from Brazilian producers, U.S. iron ore manufacturers increased efficiency and adjusted working arrangements. Competition is also strongly associated with the quality of management practices (see, e.g., Bloom et al. (2017)). Bloom et al. (2015) deploy an innovative identification strategy and show that competition improves management practices in U.K. hospitals. These papers study environments characterized by well-developed contracting institutions. Our analysis complements these findings by showing that, in a second-best world, the benefits of competition might be hampered by the presence of other market failures which are mitigated by relational contracts.[9]

More recently, Andrabi et al. (2017) and Jensen and Miller (2018) show positive effects of market competition on Pakistani schools and boat producers in Sri Lanka respectively. Unlike ours, in these two contexts buyers and sellers do not need to exchange legally unenforceable provisions of services in both directions.[10]

Finally, the paper contributes to the literature on relationships between firms when contracts are hard to enforce (see McMillan and Woodruff (1999) and Banerjee and Duflo (2000) for early contributions). A prominent context in which contracts are hard to enforce is international trade transactions (Antras (2015)). A recent literature

[8]The question of how competition affects welfare has long been regarded as central to economics (see Schumpeter (1942), Stigler (1956), Arrow (1962)).

[9]A similar tension between competition and efficiency arises in developed countries as well, e.g., in the context of innovation (Aghion et al. (2005)) and, in a way particularly related to this paper, in relational lending (Petersen and Rajan (1995)). In both cases rents are needed to attain efficiency.

[10]Ghani and Reed (2018) provides a case study of how informal trade credit arrangements between fishermen and ice suppliers evolve following the entry of an additional ice manufacturer. Our approach differs by identifying structural conditions that affects mill's ability to permanently sustain relational contracts and by exploiting cross-sectional variation across all mills in the sector.

studies relationships dynamics and quantifies the value of relationships in such contexts. For example, Antras and Foley (2015) study the evolution of trade credit terms between exporters and foreign distributors; Macchiavello and Morjaria (2015) quantify the value of relationships and identify the importance of reputational forces exploiting an exogenous supply disruption; Startz (2018) combines an original survey with an innovative structural approach to quantify search frictions and contracting costs faced by importers; Blouin and Macchiavello (forthcoming) study how relationships adapt to, and mitigate the inefficiencies caused by, strategic contractual defaults. This paper complements this agenda by asking how competition affects the sustainability of informal relationships and thus takes a step towards understanding how markets with poor contract enforcement differ from those in which contracts are enforceable.[11]

# 2 Industry Background

## 2.1 Coffee in Rwanda

*Overview:* Coffee is produced in about 50 countries around the world. Certain aspects of coffee cultivation, harvesting, processing and commercialization differ across countries. This section focuses on the Rwandan case. Coffee became widespread in Rwanda in the late 1930s following mandatory coffee-tree planting imposed by the Belgian colonial administration. At independence, in 1962, coffee represented 55% of Rwanda's exports. The decline in coffee exports started in the 1980s, accelerated with the demise of the International Coffee Agreement in 1989 and the subsequent collapse of coffee prices in the global market, and culminated 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 were around 350,000 coffee growing farmers, coffee accounted for almost 30% of the country's exports and between 12% and 15% of Rwanda's gross domestic product.

*Harvest and Processing:* The coffee cherry is the fruit of the coffee tree. 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 and its timing varies across regions depending on altitude and rainfall patterns. Coffee cherries are

[11] There has been renewed interest in interlinked transactions in agricultural chains in developing countries (see, e.g., Casaburi and Reed (2017), Casaburi and Macchiavello (2019), Casaburi and Willis (2018)). This literature typically focuses on a single aspect of interlinked transactions at a time (credit, saving, insurance). We measure bundles of complementary interlinked transactions and study how they are affected by competition.

harvested by hand, a labor intensive process requiring both care and effort. Coffee cherries, even from the same tree, do not ripen 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 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 using rocks and then dried on mats. This process is done by the farmers at home: it is cheaper but produces lower and less consistent quality.

In the wet method cherries are processed at a mill (so-called washing station or wet mill). Cherries must be taken to the mill within hours of harvest otherwise they start to ferment and rot. Mills are, therefore, scattered around the countryside. Farmers closest to the mill often take cherries to the mill's gate directly. Farmers further afield bring cherries to collection sites in which coffee collectors buy coffee.[12]

The wet method requires specific equipment and substantial quantities of water. After the cherry skin and pulp are removed with a pressing machine, cherries are sorted by immersion in water. The bean is then left to ferment for around 30 hours to remove the remaining skin. The fermentation process is carefully monitored to prevent the coffee from acquiring undesirable flavors. When fermentation is complete, the coffee is thoroughly washed with clean water. The beans are then spread out on drying tables and frequently turned by hand until completely and uniformly dry.[13]

The higher and more consistent quality wet-processed coffee, known as fully washed coffee, is reflected in prices that were, at the time of our survey, around 40% higher than dry-processed coffee both as parchment and as green coffee at the export gate. The wet method creates higher value-added along the chain (see Macchiavello and Morjaria (2015) for details).

The number of processing mills in the country increased from 3 in 2001 to around 211 in 2012 (see Figure 1). Total installed capacity in 2012 would have allowed the country to export around 70% of the harvested coffee as fully-washed (i.e. mill processed). Export data for coffee harvested in 2012, however, show that only 35% of exported coffee was fully-washed. Aggregate capacity utilization in the sector was barely above 50%. Low capacity utilization wasn't due to transitory conditions specific to the 2012 harvest and might have worsened in following years when entry of

[12] Some coffee collectors work exclusively for a mill, others operate as independent traders. Multiple collectors, often working for different mills, might source coffee from the same collection site.

[13] After the drying process is completed the coffee is hulled and consolidated for exports. Hulled coffee is referred to as green coffee. This last step is carried out by separate plants (dry mills) located around the capital city. Our analysis focuses on wet mills and farmers.

new mills expanded capacity but was not accompanied by higher export volumes of fully-washed coffee.

## 2.2 Mills Operation

*Descriptive Statistics:* To understand constraints to the operations of mills, we designed and implemented a survey of all operating mills in collaboration with the National Agricultural Exporting Board (NAEB) − the government institution in charge of the coffee sector. The survey was implemented by one of the authors towards the end of the 2012 harvest campaign (May through July). Each survey team was led by a qualified NAEB staff member. The survey covered all aspects of mills operations. The mill's manager, the main coffee collector, five randomly selected farmers and four randomly selected workers were interviewed at each mill. This paper uses data from the mill's manager and the farmers' modules.

Panel A of Table 1 reports summary statistics for mills in Rwanda. The average mill employs around 35 seasonal employees and sources from close to 400 small holders farmers. Coffee mills are thus large firms by Rwandan (Söderbom and Kamarudeen (2013)) and developing countries standards (see, e.g., Hsieh and Olken (2014)). There is dispersion in installed capacity, measured in tons of cherries per season. Installed capacity can be calculated from three aspects of the capital invested in the mill: *i*) the number of disks in the pulping machine, *ii*) the cubic metric capacity of the water tanks, and *iii*) the surface area of drying tables. Small mills have capacity of approximately 100 or 150 tons per season, most mills are medium-sized with capacity of 500 tons per season, a handful of mills have capacity in excess of 1,000 tons.

*Measuring Mills Performance:* Capacity utilization, measured as the amount of cherries processed during the harvest season divided by seasonal installed capacity, is a key driver of a mill's overall performance and is highly dispersed across mills. Leaving aside inactive mills that did not process any cherries and thus had zero capacity utilization in 2012, the median operating mill had a capacity utilization around 50%. Lack of cherries to process is not the main reason for low capacity utilization. Estimates of local production from district extension officers suggest that installed capacity is lower than local production in the areas around essentially all mills.

Capacity utilization is correlated with unit processing costs. Mills are characterized by a relatively simple technology. This facilitates the calculation of unit costs of production.[14] It takes approximately 5.5 kgs of coffee cherries to produce 1 kg of

[14] See the Appendix for details.

parchment coffee. Under a Leontieff technology approximation, the cost of producing 1 kg of parchment coffee is the sum of *i*) the price paid to farmers for the cherries; and *ii*) other operating costs, including labour, capital, procurement, transport, marketing and overheads. The former accounts for 60% to 70% of the total. After controlling for differences in unit costs explained by geographic characteristics around the mill, we find significant dispersion in overall unit costs (90/10 percentile ratio equal to 1.5) and particularly so for unit operating costs (90/10 percentile ratio equal to 2.32). Dispersion in unit processing costs is thus larger than 1.92, the corresponding figure in the typical U.S. manufacturing sector (Syverson (2011)).

## 2.3 Relational Practices between Mills and Farmers

To operate efficiently, mills rely on relationships with farmers in the surrounding areas. Panel B of Table 1 reports summary statistics for farmers from the survey. The typical farmer is a small holder who has completed primary education and owns a small coffee plantation of 500 to 1000 coffee trees. Small holder farmers in developing countries typically lack access to well-functioning input and financial markets (see, e.g., Binswanger and Rosenzweig (1986), Bardhan (1989)). Farmers resort to interlinked transactions in which a variety of services are exchanged over time with the buyers of their produce. Coffee cherries in Rwanda are no exception: transactions between mills and farmers go beyond the simple exchange of coffee cherries for cash at harvest.

The survey focused on different aspects of these transactions between mills and farmers. We refer to each aspect as a "practice". Given the lack of enforceable contracts in the rural areas of Rwanda, coffee farmers and mills must rely on informal relationships to sustain these transactions. We therefore refer to the set of practices between a mill and the supplying farmers as the *relational contract*.

Table 1 presents summary statistics for the main relational practices. We focus on those practices in which the exchange between the mill and the farmer happens over several weeks and months, i.e., those for which lack of contract enforcement matters. We distinguish between practices that are relevant before, during and post harvest. We refer to post-harvest as practices involving exchanges separate from harvest operations. For each of these practices we asked both the farmers and the manager about their use at the mill.

Before harvest, the main aspect of the relational contract is whether the mill provides farmers with inputs, extension services, and pre-harvesting loans. Gains from such practices arise from the relevant markets being poorly functioning and/or from the mill's ability to more effectively organize procurement of those inputs in bulky purchases. This type of arrangements is commonly observed in agricultural chains in developing countries, particularly in those involving large buyers sourcing from smallholders (e.g., in contract farming). Due to lack of contract enforcement, it is often difficult for the mill to ensure that, at harvest time, farmers that received inputs and loans actually deliver to the mill. Approximately 20% of the farmers report to have received inputs (e.g. fertilizers) from the mill and a similar percentage reports to have received loans. The managers survey yields similar figures on both the practices.

During harvest, the main aspect of the relational contract is whether cherries are sold on credit to the mill in exchange for part of the payment being made after the end of harvest, possibly in the form of so-called "second payment". This is beneficial for mills and farmers alike. Many mills report limited access to working capital finance to be one of the main constraints to operation. Purchasing on credit from farmers reduces mill's working capital requirements and relaxes the constraint. Many farmers report to process cherries at home to save and to be able to sell coffee when they need cash rather than at harvest. Receiving part of the payments after the end of harvest might thus help farmers with bulky expenses in the presence of such saving constraints. Since this form of trade credit is provided in-kind, input diversion is not a key concern. The main difficulty is that, due to the lack of contract enforcement, farmers might be concerned that after the end of harvest the mill might not be able, or willing, to pay the full balance still due to farmers for their deliveries. On the extensive margin, three quarters of managers and farmers report its use. On the intensive margin, however, amounts involved are often small.

Finally, as part of the relational contract, the mill and the farmers can also exchange services that are not related to harvest operations. For example, mills can help farmers with loans for bulky or unexpected expenses. Those might be related to coffee farming (e.g., help to cover the costs of replanting) or not (e.g., help with school fees). Due to lack of contract enforcement it might be difficult for mills to ensure that farmers repay those loans. On the extensive margin, 64% of farmers expect to be able to access help from the mill in case of need while 77% of mills managers report to have occasionally helped farmers with loans.

In sum we focus on the following practices: *i*) before harvest, did the farmer receive inputs and loans from the mill?, *ii*) at harvest, did the farmer sell on credit?, *iii*) post harvest, do mills help farmers with loans? We ask both farmers and managers about the use of each of the three practices at the mill. After standardizing the responses, we construct indexes for the intensity of the relationship before, during and after harvest giving equal weight to the manager's response and the average of the farmers' responses. The main variable of interest is an overall "relational" index that aggregates the three period sub-scores.

There is significant dispersion in the adoption of relational practices. Across mills, the overall relational index ranges from -2.52 to 2.01. Figure 2 shows that the use of relational practices pre-harvest, at harvest and post-harvest are positively correlated across mills. The relational index thus captures a set of complementary relational practices. Figure 3 shows that the relational index correlates positively with capacity utilization and negatively with unit processing costs. The relational index thus captures aspects of managerial practices that are appropriate in this industry.[15]

# 3 Theory

This section lays out a theoretical framework to guide the empirical analysis. The theoretical literature has already noted that competition *might* reduce parties' ability to sustain a relational contract (see, e.g., Kranton (1996) and Ghosh and Ray (1996)). Reiterating this result is thus *not* the purpose of the model. The model guides the empirical analysis in two ways. First, *when* competition reduces parties' ability to sustain a relational contract, the framework delivers predictions on how several mill and farmer-level outcomes are affected by competition. Second, competition reduces parties ability to sustain the relational contract through separate channels. The model offers guidance on how to empirically separate these channels.

First, we present the set-up of the model, illustrate preferences, technology and interlinked transactions between mills and farmers when contracts are perfectly enforceable. We then remove perfect contract enforcement and study a monopsonist mill that sources coffee from farmers through spot transactions alone. We then explore the conditions under which a relational contract between the mill and the farmers is sustainable. Finally, we introduce competition and derive the testable predictions.

[15]We abstract from sources of opportunism that occur over very short periods and for which lack of contract enforcement does not matter. For example, farmers are sometimes paid a few days after delivery. This short-term trade credit arises from liquidity management with mills holding cash for payments only few days a week. A mill could, in principle, default on this short-term credit. Similarly, a mill could hold-up farmers at the mill's gate arguing over the price or the quality of the cherries. While, anecdotally, this behavior is known to happen it is not widespread: it would be difficult for a mill to source coffee for the rest of the season following such behaviour. Extensive conversations in the field and the empirical analysis confirm this interpretation.

### 3.1 Set-Up

*Players and Preferences*: A risk-neutral 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 identical 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). Farmers derive utility from consumption at harvest, $c_1$, and post-harvest, $c_2$, with preferences given by $u(c_1, c_2) = \min\{c_1, c_2\}$. These preferences capture farmers' demand for within-season consumption smoothing.[16] All parties have a discount factor $\delta < 1$ across seasons. For simplicity, there is no discount within season.

*Production, Technology and Timing of Events*: Figure 4 illustrates the timing of events during a season. At pre-harvest the mill can provide inputs (e.g., fertilizers, loans) to the farmers at cost $\kappa$ per farmer. The farmer must exert effort, at cost $e$, to use the inputs correctly. If she does so, the quality of mill-processed coffee is increased and its value enhanced by a factor $\lambda$.

Production and processing take place at harvest. A quantity $q$ of coffee cherries becomes available, and must be processed, at harvest. Once processed, coffee becomes storable. Two technologies are available: home processing and mill processing. Both technologies yield one unit of output per unit of cherries. Home processing is performed by the farmer at home and, for simplicity, we assume it entails no additional cost. 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 an exogenous installed capacity, $\mathbf{C}$. The mill, a price taker in all other markets except coffee cherries, incurs additional processing unit costs, $c(Q) > 0$, with $c'(Q) \leq 0$, when $Q \leq \mathbf{C}$, and $c(Q) = \infty$ otherwise.

The farmer can sell home processed coffee at an exogenous price $\rho$ at harvest and post-harvest. Mill processed coffee sells for $v > \rho$. We make the following assumption:

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

The assumption $v > \rho + c(0)$ implies that mill processing is efficient. The assumption $v > \frac{\kappa+e}{\lambda q}$ implies that the mill and the farmers' costly actions at pre-harvest are efficient. The assumption $2\frac{\kappa+e}{\lambda q} > v$ avoids a taxonomy of uninteresting cases.

[16] The functional form of the utility function can be relaxed at the cost of more tedious algebra without altering the main insights of the analysis.

### 3.2 Benchmark Case: Perfect Contract Enforcement

When contracts are perfectly enforceable actions are contractible and parties can commit to promises. 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 farmer $i$. 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, $P_2^i$. This second payment corresponds to the mill repaying trade credit received by the farmer at harvest. These contractual terms thus capture the pre-harvest and harvest sub-scores of the relational index introduced in the previous section. For simplicity, the theoretical analysis abstracts from loans and help to farmers unrelated to harvest operations (post-harvest subscore in the empirical measure). It is easy to add those to the model without affecting any of the key results.

Each farmer $i$ independently decides whether to accept or reject the contract. If the farmer rejects the contract, she harvests quantity $q$, processes it at home, and sells it on the market, deciding between harvest and post-harvest sales to maximize her utility. 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 and the value of this choice by $u^i = \rho q/2$.[17] If the farmer accepts the mill's offer, the contract must then be respected by all parties. Denote by $u_{\mathcal{C}}^i$ the utility of farmer $i$ from accepting contract $\mathcal{C}$. We focus on a symmetric solution in which the mill offers contract $\mathcal{C}$ solving

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

By Assumption 1 the optimal contract involves no home processing (i.e., $q_m = q$), $\mathbf{I}_k = 1$, and $\mathbf{I}_e = 1$. 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 setting $P_1 = P_2 = P$ such that $P = \rho q/2 + e$.

**Observation 1**: *Under perfect contract enforcement: 1) the mill provides inputs to farmers and farmers exert effort, 2) farmers only sell to the mill, 3) the mill pays both a spot price and a second payment.*

[17] 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 $\min\{c_1, c_2\} = \rho q/2$.

### 3.3 Monopoly Mill under No Contract Enforcement

We now turn to the case in which contracts between the mill and the farmers are not enforceable. More specifically (see discussion in footnote 15) we assume that contracts are enforceable within a sub-period but not across sub-periods. The mill posts a unit price for the harvest season, $p_1$, (to which it can commit) and promises an additional second payment, $P_2$, to be paid post-harvest. Given posted prices and beliefs, farmers decide 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). The mill finally decides whether to pay any promised $P_2$.

The model has to be solved by backward induction. The mill always defaults on any promised $P_2$. Farmers, therefore, base their decision only on posted prices $p_1$ at harvest time. Given $p_1$, farmers equate consumption at harvest ($c_1 = p_1 q_m + \rho q_{\rho 1}$) and post-harvest ($c_2 = (q - q_m - q_{\rho 1})\,\rho$) subject to the constraint $q_m + q_{\rho 1} + q_{\rho 1} \leq q$. A farmer's supply curve is then given by[18]

$$q_m(p_1) = \begin{cases} 0 & \text{if } p_m \leq \rho \\ \frac{\rho}{p_1+\rho} q & \text{otherwise} \end{cases} \tag{2}$$

The mill, taking as given the supply curve of each farmer $q_m(p_1)$, sets $p_1$ to maximize profits. 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 farmers and farmers do not exert effort, 2) farmers sell only a fraction of their produce to the mill and home process the rest, 3) the mill does not pay any second payment.*

### 3.4 Relational Contracts

We now consider relational contracts between the mill and the farmers. A relational contract is a plan $\mathcal{C}^R = \{\mathbf{I}_k^t, \mathbf{I}_e^t, q_m^t, P_1^t, P_2^t\}_{t=0,1,\ldots}^{\infty}$ 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 up the

[18]The farmer's supply curve is *backward* bending for $p_1 > \rho$. Our results rests on farmers having a demand for saving, not on this feature of farmers' supply curve. Casaburi and Macchiavello (2019) provides experimental evidence consistent with a backward bending supply curve arising from saving constraints in the Kenya dairy sector.

relationship and to obtain their outside options forever following any deviation. We derive conditions under which a stationary relational contract in which *i*) pre-harvest $\mathbf{I}_k^t = 1$ 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; can be sustained. We assume a multilateral relational contract in which if the mill reneges against *any* farmer then farmers jointly punish the mill as in Levin (2003). The assumption can be relaxed without altering the key insights.

Denote with $\pi_r$ and $\pi_s$ the seasonal profits of the mill under the relational contract and under spot transactions respectively. The incentive compatibility constraint of the mill is

$$\pi_r - \pi_s \geq (1-\delta) \max\left\{\frac{P_2}{\delta}, \kappa\right\}. \tag{3}$$

We focus on the case $\frac{P_2}{\delta} > \kappa$ and avoid a taxonomy of cases that yield no further inisight. Denote with $u_r$ and $u_s$ the seasonal monetary payoff of the farmer under the relational contract and under spot transactions respectively. The incentive compatibility constraint of the farmer is

$$u_r - u_s \geq e. \tag{4}$$

Two remarks on (4). This constraint is needed to induce the farmer to exert effort. The incentive constraint needed to incentivize the farmer to sell is given by $u_r - u_s \geq \delta e$ and is thus never binding if (4) is satisfied. Second, the farmer is indifferent between the relational contract and spot transactions. This is a special feature of this model. Under more general scenarios, farmers must be given rents to comply with, and are thus better off under, the relational contract. We discuss this aspect further below.

The mill offers a relational contract that maximizes profits subject to (3) and (4). As in the previous section, $u_s = \rho q/2$. Substituting the binding constraint (4) into (3) the problem of the mill can be rewritten as finding the maximum quantity $q_r$ subject to

$$\frac{\delta}{1-\delta}\left[(1+\lambda)vq_r - c(q_r) - e - \kappa - (vq/2 - c(q/2))\right] \geq P_2. \tag{5}$$

By standard logic a solution exists if $\delta$ is large enough. 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.*[19]

[19] Assumption 1 implies that if a relational contract can be sustained at all, it entails $q_r = q$. In general however relational contracts might be feasible even without achieving the first best. We focus

### 3.5 Relational Contracts Under Competition

The effect of competition on parties' ability to sustain relational contracts has been theoretically studied elsewhere (see, e.g., Kranton (1996) and Ghosh and Ray (1996)). We focus on deriving additional predictions and understanding the mechanisms through which competition alters the ability to sustain relational contracts in our context. Consider the case in which a competing mill locates near the existing monopolist. For simplicity, assume that the competing mill offers spot transactions in a particular season. The logic is of course strengthened if (farmers believe) the competing mill offers spot transactions in future seasons as well.[20]

For competition to alter the conditions under which a relational contract is sustainable, it has to be that the farmer's side-selling constraint becomes the binding constraint. Denote with $\mathbf{p}_0$ the price offered by the competing mill. If the farmer accepts the offer, she sells $\mathbf{q}_0$ to the competing mill and set $\mathbf{p}_0\mathbf{q}_0 = (q - \mathbf{q}_0)\,\rho$. The no side-selling constraint for the farmer is given by:[21]

$$P_r + \frac{\delta}{1-\delta}u_r \geq \frac{q\rho\mathbf{p}_0}{(\mathbf{p}_0+\rho)} + \frac{\delta}{1-\delta}u_s. \tag{6}$$

Substituting for the corresponding values of $u_r$ and $u_s$ the binding constraint (6) gives the minimum transfer the mill must pay to prevent competition, $P_2^c = (1-\delta)\,\frac{q\rho\mathbf{p}_0}{(\mathbf{p}_0+\rho)} + \delta\,(\rho q/2 + e)$. The incentive constraint of the mill then becomes

$$\frac{\delta}{1-\delta}\left[(1+\lambda)vq_r^c - c(q_r^c) - e - \kappa - (vq/2 - c(q/2))\right] \geq P_2^c. \tag{7}$$

Since $P_2^c > P_2$ the mill incentive compatibility constraint under competition (7) is, thus, harder to satisfy than (5). This implies:

**Observation 4**: *There exists a 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 monopsony but is not sustainable under competition. When this happens, transactions under competition occur as described in the no contract enforcement case.*

---

on this special case to ease exposition without altering any of the key insights. See Appendix for details.

[20] The relational contract is even harder to sustain if the competing mill enters offering a (credible) relational contract. In such cases, however, competition leads to a reallocation of relationships, but not to the relationship breakdown. We focus on deriving predictions when competition leads to relationship breakdown.

[21] Note that the farmer's side-selling constraint can become the binding one even when the competing mill's willingness to pay is *lower* than that of the incumbent mill: the *threat* of competition alone can bite on the relational contract.

### 3.6 Predictions

*Relational Practices are Complementary:* Competition *might* destroy the relational contract between the mill and the farmers. *When* this happens, the first prediction of the model is that *all* the relational practices are affected. That is, although competition directly affects only farmers' selling opportunities at harvest, the mill's incentive compatibility constraint bundles practices together. The collapse in the incentive constraint reduces parties ability to sustain pre- and post-harvest practices as well.

*Additional Outcomes:* The model predicts that when competition destroys the relational contract, *production*, *costs*, *prices* and *farmers' welfare* are all affected in a coherent way. That is, the model delivers a cluster of predictions on additional outcomes. First, when the relational contract collapses there is no market price at which farmers sell *all* their production for processing 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's efficiency are lower than under a relational monopoly. Output quality suffers too, as mills no longer provide inputs before harvest and farmers do not exert adequate effort.

Second, the effect of competition on prices paid to farmers 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, however, when the relational contract collapses, mills and farmers no longer invest in pre-harvest inputs and effort. Since the relational contract compensates farmers for effort costs the net present value of payments to farmers is unambiguously higher under the relational contract. Prices paid during harvest might also fall as a result of competition.

Finally, the effect of competition on farmers welfare is also ambiguous. In the framework above competition makes farmers weakly better off. As already noted, this happens because under the relational contract the participation constraint binds and farmers earn no rents. In general, however, lack of contract enforcement implies rents for farmers. For example, farmers earn rents if we relax the assumption of perfect observability of effort. In this case the incentive constraints implies that the farmer's participation constraint is slack. Alternatively, farmers earn rents if the mill is not able to perfectly discriminate the relational contract offered to heterogenous farmers. In this case, only the participation constraint of the *marginal* farmer binds and inframarginal farmers earn rents. Under either scenarios, competition reduces the welfare of at least some farmers.

*Mechanisms:* The incentive constraint (7) also reveals *two* distinct mechanisms through which competition erodes the mill's ability to sustain a relational contract. First, there is a *direct temptation* mechanism operating through the right hand side of (7): it is harder to commit to the higher transfer that must be promised to the farmer. This mechanism is direct in the sense that the mill would find it more difficult to sustain the relational contract with those farmers for which competition yields alternative selling opportunities.

Second, there is an *indirect profit* mechanism operating through the left hand side of (7). By reducing sourcing from farmers directly affected, competition reduces the mill's aggregate profits and, indirectly, the mill's ability to sustain the relational contract even with those farmers that are not directly affected by competition.

Empirically we attempt to separate the two mechanisms as follows. From the perspective of an individual farmer we can distinguish the effect of competition from mills that are located near the farmer *and* the mill (farmer-competition) from the effect of competition from mills that are competing with the mill but are otherwise far from the farmer (mill-competition). Farmer-competition affects the relational contract with the farmer through both the direct and indirect effects. Mill-competition affects the relational contract with the farmer only through the indirect effect. The effect of mill-competition thus identifies whether the indirect *profit* mechanism operates. The difference between the effects of farmer-competition and mill-competition identifies whether the direct *temptation* mechanism is at work.

***Summary of Predictions:***

A. *Competition might reduce relational contracts between farmers and mills.*

B. *When this happens, the following is observed:*

   1) *All relational contract practices are undermined: inputs/loan pre-harvest, trade credit and second payments, and help/assistance to farmers*;

   2) *Mills have lower capacity utilization, higher processing costs and produce lower quality*;

   3) *Prices paid to farmers at harvest might increase or not, but farmers sell fewer cherries at harvest to any mill and are worse off*;

C. *Competition at the mill-level reduces relational contracts with farmers that can directly sell to the competing mill as well as, indirectly, with those farmers that cannot, albeit to a lesser extent.*

# 4 Does Competition Break Relational Contracts?

This section asks whether competition breaks relational contracts (prediction **A**)? We begin by describing the baseline measure of competition and presenting OLS estimates. The OLS estimates could be biased either due to omitted variables (e.g., if there are unobservables factors that correlate with both competition and with either the desirability or the feasibility of relational contracts); reverse causality (competitors might locate near mills with either worse or better relational practices); and measurement error (it is difficult to measure which mills actually compete with each other). Given these concerns we lay out an instrumental variable (IV) approach. We present the main IV results and a battery of robustness checks. Our results indicate that competition reduces the incidence of relational contracts between mills and farmers. When this happens, the model yields predictions about additional outcomes and about the mechanisms at play. The next section explores the consequences of relational contracts breakdown (predictions **B**) while section 6 tests for mechanisms (prediction **C**) and further discusses the exclusion restriction of the IV approach.

## 4.1 Measuring Competition

The baseline measure of competition is the number of mills within a 10 km radius from the mill. We say that two mills compete with each other if their catchment areas overlap. Defining competition faced by a given mill thus requires a definition of the mill's catchment area. There are two alternative approaches to define the catchment area. A first approach is to directly ask mills' managers about the size of what they consider to be the mill's catchment area. A second approach, instead, uses the typical density of trees to estimate the size of the area such that, if the mill processed half of the coffee cherries in that area, it would operate near full capacity. For mills with average capacity both approaches give a radius of 4.0 to 4.5 km (measured in Euclidian distance). We take a conservative approach that stretches the catchment area to a 5 km radius. Given our definition, then, a mill potentially competes with other mills located within a 10 km radius (see Figure A1 for an illustration).

Figure 5 illustrates the distribution of the number of mills within a 10 km radius from each mill. There is significant dispersion in the intensity of competition faced by mills. While there are quite a few isolated mills, the average mill has 6 competitors. We can use the survey to check whether our measure of competition captures the degree of competition actually experienced by the mill's managers. The survey asked the mill's manager the number of other mills that source coffee cherries inside the mill's catchment area and at the nearest collection site used by the mill. The average manager reported competition from about 2 other mills at the nearest collection site and from 6 within the catchment area. The correlation coefficient between the survey measure and our baseline measure is 0.77 and highly significant. The baseline measure thus captures well the intensity of competition actually experienced by mills.

The baseline measure takes a one-size-fits-all approach to define competition. Mills are, however, highly heterogeneous suggesting that a mill-specific measure of competition might be better suited for our analysis. The reason we prefer our baseline approach is that mill's specific conditions (e.g., installed capacity) might endogenously respond to both competition and to mill's practices. Mill's specific measures of competition thus introduce additional sources of bias. The baseline measure avoids that. To the extent that the baseline measure suffers from measurement error, OLS results will be biased towards zero. For expositional simplicity, we present OLS results using only the baseline measure and discuss several robustness checks after presenting the IV analysis.

## 4.2 Competition and Relational Contracts (Prediction A): OLS

Denote with $RC_m$ the relational index at mill $m$ and with $C_m$ the number of mills within 10 km of mill $m$. The OLS specification is given by

$$RC_m = \alpha + \beta C_m + \eta X_m + \gamma Z_m + \varepsilon_m, \tag{8}$$

where $X_m$ and $Z_m$ are vectors of controls at the mill level ($m$) and $\varepsilon_m$ is an error term. The vector $X_m$ includes mill's characteristics (age, NGO-support and cooperative status). The vector $Z_m$ includes geographic controls for potential drivers of the mill's performance within the mill's catchment area: elevation, slope, x- and y-coordinates, historical suitability for coffee, density of coffee trees, length of roads and rivers.

Column 1 in Table 2 reports the results. Competition negatively correlates with relational contracts: an additional competing mill is correlated with a 0.112 standard deviation lower relational index. The OLS estimates, however, might be biased due to a number of concerns and cannot be interpreted as conclusive evidence of a negative impact of competition on relational contracts. For example, unobserved local conditions, such as farmers' skills or entrepreneurial attitude, might both be conducive to establish relational contracts and attract more competition in the area. In this case the OLS coefficient is upwardly biased. Conversely, better access to inputs and/or financial services could attract competition to the area but reduce farmers' demand for relational contracts. Potential entrants might also locate next to poorly run mills that score badly on relational contracts practices. In such cases, the OLS coefficient is biased downward. Furthermore, as noted above, the one-size-fits-all approach in our baseline measure of competition introduces measurement error which could bias the OLS estimate towards zero.

### 4.3 Construction of the Instrument: Entry Model

Given these concerns we 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 (in particular the use of relational contracts with farmers) other than through its effect on competition (exclusion restriction). To construct our instrument we combine *i*) the spatial nature of competition embedded in the notion of catchment area defined above with *ii*) drivers of suitability for mill placement (henceforth, "suitability"). We derive those drivers from an engineering model for optimal mills' placement in Rwanda. Conditional on suitability *within* the mill's catchment area, competition is instrumented with suitability in the region *around* the mill's catchment area. Given our baseline definition of catchment area, the instrument for competition is then given by suitability for mill placement between 5 and 10 km from the mill, conditional on suitability (and other controls) within the 5 km radius catchment area.

We take advantage of an engineering model to construct our measure of suitability. In the early 2000s, when only a handful of mills were operating in Rwanda, a program coordinated by USAID involving engineers, agronomists and soil specialists developed an engineering model for the optimal placement of mills in Rwanda (see, Schilling and McConnell (2006)). Given the particularly rugged nature of Rwanda, the model intended to identify suitable sites for mills construction at a high resolution level taking into account a vector of characteristics to be then aggregated into a suitability score. The model, however, was never implemented because the required GIS high resolution ortho-photos were not available for the whole of Rwanda at the time. Subsequent entry of mills was thus not restricted nor limited to locations satisfying the engineering model's criteria. We assembled all the data required ex novo and are thus able to implement the engineering model for the first time. Using remote sensing and GIS tools on ortho-photos at the $25m^2$ resolution we run the engineering model for the whole of Rwanda at a resolution of $1km^2$.

The engineering model specified four criteria for a mill placement: *#1)* Outside National Parks, Reserves and other protected areas; *#2)* in sectors with at least 30,000 coffee trees;[22] *#3)* Within 3 km from a spring source, at an elevation between minus 10 and minus 30 meters from the spring; *#4)* Within 1 km of a road. For each $1km^2$ square in Rwanda (henceforth, "cell") we define dummies for whether it satisfies each of these four criteria or not. We utilize these dummy variables and their interactions to predict the actual placement of mills at the cell level. There are thousands of potential cells where mills could have entered and 211 in which a mill had entered by 2012. For each cell we obtain a score summarizing the suitability of that particular location based on the engineering model. Finally, the predicted scores are aggregated at the mill level, taking averages of the scores in the cells within the mill's catchment area and the area around it as defined above.

Figure A2(A) illustrates spatial variation in the engineering model's criteria. The dark grey cell boxes are ineligible for entry due to their land cover. The lightest green illustrates cell boxes that satisfy the number of trees necessary for entry, the brightest green areas highlight where the cell 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. Grids not satisfying these two criteria are assigned a score equal to zero.[23] Within the sample of cells satisfying criteria #1 and #2 we take into account the remaining criteria #3 and #4 by running a probit model that predicts whether the cell has a mill or not. The specification includes polynomials in distances to springs and roads, elevation, average slope in the cell, density of coffee trees (or suitability for coffee), longitude, latitude and the interactions of these variables. Table A1 in the Appendix reports the results. The estimates lend support to the engineering model. Columns 1 to 3 show that both criteria #3 and #4 in isolation predict mill's placement. Column 4 shows that the interaction between the two criteria predicts mill's placement,

[22]Sectors are the third-level administrative units in Rwanda, with an area of approximately 50 $km^2$.

[23]Criterion #2 requires a cell to be placed in a sector with enough coffee trees. A potential concern with the inclusion of this criteria is that farmers might endogenously adjust planting decisions in response to competition and/or relational practices. Note, however, that conditional on being in a sector with a sufficient number of trees (all mills are) the criterion (and, therefore, the instrument) does not correlate with the density of trees in the area. We can either include or exclude density around the mill as a control as a further robustness check. Furthermore, farmer level results (to follow) show that competition does not affect farmers planting decisions.

consistently with the engineering model. We use estimates in Column 4 to predict a suitability score for each $km^2$ cell. Figure A2(B) illustrates the predicted score from the model. The instrument is given by the average suitability score within 5 to 10 km area from the mill (akin to a cross-section surface of a donut) conditional on the average of the score and its individual components within the 5 km catchment area.

### 4.4 Competition and Relational Contracts (Prediction A): IV

We instrument for competition using the average predicted score from the engineering model in the donut area between 5 and 10 km from the mill. Specifically, the first stage is given by

$$C_m = \alpha + \widehat{\beta} S_m^{5/10} + \beta S_m^{0/5} + \widehat{\gamma}_0 X_m + \widehat{\gamma} Z_m + \mu_m \tag{9}$$

where $S_m^{5/10}$ is the average predicted score in the donut area between 5 and 10 km from mill $m$, $S_m^{0/5}$ is the score inside the mill's catchment area and $C_m$ is the number of mills within 10 km from mill $m$. The vectors $X_m$ and $Z_m$ are mill controls described in equation (8). The exclusion restriction is satisfied if, conditional on suitability conditions within the mill's catchment area, average suitability in the 5-10 km area only affects a mill's operation through its effect on competition. Potential threats to the exclusion restriction are discussed in Section 6.2.

Figure 6 shows a strong first stage: the predicted score $S_m^{5/10}$ strongly correlates with competition $C_m$. The top-left panel shows the raw correlation between the instrument and competition, the top-right panel includes the vector of controls in the model, $S_m^{0/5}$, $X_m$ and $Z_m$. Column 2 in Table 2 reports the results. An increase of one standard deviation in the instrument $S_m^{5/10}$ is associated with mill $m$ facing competition from 2.072 additional mills (p-value $< 0.01$, $R^2 = 0.76$).

Figure 6 also shows (bottom panel) a strong reduced form relationship between the instrument, $S_m^{5/10}$, and the relational index, $RC_m$. Column 3 in Table 2 reports the estimates. An increase of one standard deviation in the instrument $S_m^{5/10}$ is associated with a reduction of 0.527 standard deviations in the relational index at mill $m$ (p-value $< 0.01$).

Column 4 in Table 2 reports the 2SLS estimates. An additional mill within a 10 km radius from the mill causes a reduction of 0.254 standard deviations in the relational index. The effect is economically sizeable. The comparison between the IV estimates in Column 4 and the OLS estimates in Column 1 reveals that the IV estimates are about twice as large as the OLS (-0.254 versus -0.112). 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 the use of relational contracts.[24]

The specification assumes a linear effect of the number of competing mills on the relational index. The model, however, suggests that the relationship might be non-linear: relational contracts break down only when there is competition beyond a certain threshold. Aggregating over mills with heterogeneous threshold we thus expect the negative effect of competition to become stronger as competition intensifies, at least up to a certain point. Figure A3 in the Appendix explores the functional form of the relationship between competition and relational contracts. The Figure reports results from Hall and Horowitz (2005) non-parametric IV estimation. Consistent with the model's logic, the estimates exhibit a decreasing and concave relationship between relational contracts and competition over the entire range of observed competition levels. The slope is relatively flatter for competition from fewer than 4 mills and then becomes steeper once competition intensifies.

## 4.5 Robustness of IV Estimates

*Inference:* The analysis is robust to different assumptions on the structure of the error term in equation (8). There are two main concerns: *i*) the instrument is the average of cell suitability score, a predicted variable; *ii*) errors could be spatially correlated across mills. Table 2 reports confidence interval under four different assumptions on the structure of the error term. Considering the baseline IV specification in Column 4, we describe each procedure starting with the most conservative. A first procedure simply bootstraps the two-stage estimation, using the entire sample of cell points to predict the score at the cell level for the mill-level regression. This procedure yields the most conservative standard error for the main coefficient of interest (0.065, p-value $< 0.01$). The following two procedures allow for error terms to be spatially correlated across mills. A first procedure allows for arbitrary spatial clustering as in Konig et al. (2017). A second procedure adjusts for spatial clustering as in Conley (1999). These two procedures yield standard error estimates of 0.058 and 0.051 respectively. Finally, we implemented subsampling bootstrap, in which we randomly subsample 90% of cell points in the index construction stage and then reestimate the engineering model, scores, first and second stages. This procedure delivers a standard error estimate of 0.043. To simplify exposition, the reminder of the analysis only reports estimates with

[24]For ease of comparison between the OLS and IV estimates, Column 1 in Table 2 already includes the average suitability score within 5 km as a control.

the most conservative method.

*Catchment Area, Instruments and Competition:* Additional unreported results explore the robustness of our analysis to alternative definitions of the mill's catchment area, of competition and instruments. First, results are robust if we consider catchment areas of different size. We can consider catchment areas with smaller (larger) radius of 3.5 (7) km and associated donut area in between 3.5 and 7 (7 and 15) km. Estimates yield a coefficient which is about one and a half (two thirds) of the baseline coefficient, always statistically significant. In each case, the IV identifies a weighted average of the effects of additional competitors within the relevant area. The decreasing magnitude identified as we extend the size of the catchment area is consistent with the impact of an additional competitor decreasing with the distance of said competitor from the mill. Competitors beyond 10 km do not affect relational contracts. The baseline definition of catchment area is preferred as it strikes a good compromise between two conflicting goals. On the one hand, a wider area includes mills that are not actual competitors. On the other hand, a narrower area is more likely to violate the exclusion restriction.

Results are also robust when we allow the assumed size of the catchment area to vary with mill's characteristics. We can construct a mill specific catchment area by using distance from the mill along roads, thus allowing the effective size to depend on the density of roads around the mill. Alternatively, we can determine the catchment area of the mill using the surrounding density of trees relative to capacity, thus allowing the effective size to depend on the mill's capacity and farmers' planting decisions. Results are again, qualitatively consistent with the baseline estimates. We prefer the baseline definition of catchment area as it does not directly depend on mills' and farmers' decisions.

Finally, we can explore the robustness of the findings using either alternative instruments or alternative definitions of competition. First, we can use the average of each score criterion and their interactions as separate instruments, or we can predict suitability with a liner model instead of a probit. These alternatives yield nearly identical and precisely estimated results. Second, we can measure competition as directly reported by the manager, or aggregating the capacity of surrounding mills, or as distance to the n$^{th}$ nearest competitor. These specifications yield qualitatively similar estimates, albeit sometimes coefficients are less precisely estimated.

# 5 The Consequences of Relational Contracts Breakdown

The previous Section shows that competition decreases the use of relational contracts. When this happens, the model delivers a cluster of additional predictions about how relational practices move together and about both mill-level and farmer-level outcomes. This section tests these additional predictions.

## 5.1 Complementarities in Practices (Prediction B1)

The first implication of the model is that relational contracting practices are *complementary*. Competition alters directly only farmers ability to sell cherries at harvest to a competing mill. When the incentive constraint is violated, however, *all* practices become unsustainable. Competition thus undermines *all* aspects of the relational contract. Table 3 reports both IV (Panel A) and OLS (Panel B) specifications considering the relational practices one at a time. For each relational practice, the Table reports both specifications using farmers' answers, managers' answers, and the aggregate of the two.[25]

Columns 1 to 3 ask whether competition reduces relational practices in which the mill provides inputs and loans to farmers before harvest. Regardless of whether we ask farmers or managers, competition causes a reduction in the use of this practice. When answers from farmers and managers are aggregated, we find that competition from an additional mill reduces the use of this practice by 0.171 standard deviations.

Columns 4 to 6 ask whether competition reduces the relational practice through which the mill sources cherries on credit at harvest. Regardless of whether we ask farmers or managers, competition causes a reduction in the use of this practice. When answers from farmers and managers are aggregated competition from an additional mill reduces the use of this practice by 0.190 standard deviations.

Finally, Columns 7 to 9 ask whether competition reduces the use of the relational practice in which the mill provides assistance and help to farmers post-harvest. Competition from an additional mill reduces the use of this practice by 0.191 standard deviations.[26] Columns 10 and 11 aggregate the three relational contract practices by respondent type. We find virtually identical estimates of the effect of competition on the use of relational practices as reported by farmers and managers. Note also that for all practices the IV estimate is larger than the OLS.

[25] The farmer-level specifications include farmer's age, gender, place of birth, education level, cognitive skills, distance from the mill and farm's coffee tree holdings as additional controls.

[26] The p-value is 0.12 when the practice is reported by the manager.

The model focuses on relational practices for which lack of contract enforcement matters. Those are practices for which the mill and the farmer rely on non-enforceable promises across several weeks (pre, during and post harvest). Lack of contract enforcement is not a concern for sources of opportunism that occur over very short periods, such as the mill defaulting on cash payments. Accordingly, Column 13 in Table 3 reports a placebo specification in which we consider short-term credit and advances during harvest as dependent variable. We argued above that, since the vast majority of payments to farmers are made in cash this practice is driven by liquidity management and is thus not part of the relational contract between the mill and the farmer. The results confirm that competition does not impact this type of short-term credit between the mill and the farmers.

The evidence thus supports the predictions of the model. Competition reduces the use of *all* relational contracts by a magnitude that is essentially identical across practices.

## 5.2 Mill Outcomes: Operations and Quality (Prediction B2)

*Operating Costs:* The model predicts that a breakdown in the relational contract with farmers is associated with changes in mill's outcomes. Table 4 investigates these predictions. Columns 1 to 4 explore unit costs. Column 1 shows that unit costs increase by 3.7% with an additional competing mill. Column 2 shows no effect on prices paid to farmers during harvest. Column 3 presents a placebo: competition has no effect on the conversion ratio from coffee cherries to processed parchment, a parameter of the production function. The combination of Columns 2 and 3 implies that competition has no effect on the cost of cherries. The cost of cherries accounts for about 60% of the overall unit costs at the typical mill. The coefficient in Column 1 must thus be explained by increases in other operating costs. Accordingly, Column 4 shows that an additional competing mill increases processing unit costs by approximately 7%.

The increase in unit costs arises from both lower and more sporadic deliveries. Column 5 shows that competition reduces capacity utilization, computed as the ratio of the total amount of cherries processed during the harvest season divided by the total capacity installed. An additional competing mill is associated with 4.6% lower capacity utilization (p-value 0.12). With average capacity utilization in the industry around 50% this is a sizeble effect. This effect is conditional on mills operating during the harvest season. On the extensive margin competition also increases the likelihood the mill does not operate at all for the whole season.

The breakdown in relational contracts with farmers makes deliveries harder to plan for. Column 6 shows that competition does not affect the number of weeks the mill is in operation during harvest. Column 7 shows that competition, however, increases the likelihood the manager reports to have had days with too few cherries to process (p-value $<0.15$). In fact, competition increases the likelihood that the manager reports to have had *both* days with too many and too few workers at the mill (Columns 9 and 10). The difficulty in planning results in higher labour costs. Column 8 shows that the labour component of unit costs increases with competition: an additional mills increases unit labour costs by nearly 15%.

Irregular deliveries from farmers, however, increase labour unit costs only if the mill cannot perfectly adjust hired labour to daily cherries availability. Additional survey evidence confirms this to be the case. While 65% of mills do revise employment plans weekly depending on cherry procurement and market conditions, hiring is not perfectly flexible and arrangements between mills and workers also include elements of relational contracting. The majority of seasonal workers is paid weekly, bi-weekly or monthly, rather than daily. Firms thus do not turn down workers when there are not enough cherries arriving at the mill. For example, 73% of mill managers report that they would turn down only some, and 12% report none, of the workers if there were very little cherries to process. Note that mills competing over cherries do not compete for workers and therefore competition does not increase unit labour costs by raising workers wages. Mills are located in densely populated rural areas with excess labour supply. Unreported results show that competition has no impact on wage rates nor on the likelihood the manager reports difficulties in hiring workers.[27]

*Product Quality:* The model also predicts that when relational contracts break-down, the quality of the coffee produced by the mill suffers. This happens because the mill does not provide inputs to farmers and farmers do not exert appropriate effort. In particular, farmers harvest less frequently and end up mixing cherries that are ripe with others that are either too ripe or not ready yet.

To test this prediction we collected random samples of processed coffee from each mill. Each sample was inspected and "cupped" at the national coffee board's laboratory in Kigali under the supervision of one of the authors. The cupping process scores each sample along several dimensions of quality related to both physical characteristics of

[27] Besides labour, other processing costs are given by capital, transport, procurement and other types of costs. Table A2 in the Appendix provides some evidence that competition makes it more difficult for mills to access working capital. Although the model does not consider this mechanism explicitly, this is consistent with competition reducing trade credit received from farmers. We do not find significant effects of competition on transport, procurement and other costs.

the processed beans as well as defects that emerge following the roasting process. Physical characteristics and defects can be classified depending on their most likely origin: plant genetics, farmer's husbandry practices and mill processing.

Table 5 presents the results. Column 1 shows that competition decreases the overall quality score of coffee processed by the mill. An additional competing mill reduces the quality score by 0.15 standard deviations. Columns 2 to 4 separate the score into different quality components depending on whether they are mostly under the control of the farmer (column 2), mill (column 3) or are genetically predetermined (column 4). We construct an index that captures aspects of quality that are under the direct control of farmers. The index aggregates two dimensions of quality: bean size and pest damages. Given planted variety, smaller bean size is a consequence of poor harvesting practices. Severe insect and pest damages arise from inadequate use of insecticides at the farmer level. Column 2 shows that an additional competing mill decreases the index of farmer-related quality by 0.171 standard deviations.

We also construct an index that captures aspects of quality that are under the direct control of mills. The index aggregates moisture content, floating beans and broken beans as dimensions of quality. Those quality dimensions are mostly influenced by sorting and drying practices at the mill. Column 3 shows a smaller impact of competition on the index of mill-related practices, an additional competing mill decreases the index of mill-related quality by 0.1 standard deviations. The effect is not only smaller, but is also not statistically significant at conventional level. Column 4 shows that competition has no impact on a dimension of quality directly related to the genetic variety of coffee grown by the farmer.

The evidence thus supports the model's predictions that competition increases mills' operating costs and reduces the quality of the coffee produced by the mills through its negative impact on relational practices with farmers.

### 5.3 Farmer Outcomes (Prediction B3)

The model predicts that a breakdown in the relational contract with the mill is also associated with changes in several farmer-level outcomes. In particular, 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 cannot rely on mill's second payments to smooth cash flows), *iii*) a reduction in access to inputs. Furthermore, simple extensions of the model that allow for either heterogenous farmers or for farmers' moral hazard imply that a breakdown in the relational contract lowers welfare for at least some farmers.

Table 6 tests these predictions with farmer level specifications. Column 1 essentially confirms the finding of Column 2 in Table 4: competition has no effect on prices received by farmers. While the detected effect is positive and statistically different from zero, it is very small. An additional mill increases prices reported by farmers by less than 1%. Note that this is the price farmers report for their sales of cherries during harvest. Since competition reduces payments made to farmers after the end of harvest this estimate likely provides an upper bound to the effect of competition on the net-present-value of payments to farmers.

Column 2 shows that competition reduces the share of a farmer's production sold as cherries during harvest. That is, competition between mills actually *increases* the share of coffee that is home processed. Column 3 also shows that competition increases the likelihood that farmers report saving as the main motivation for processing coffee at home rather than selling cherries at harvest. Taken together, these two results confirm the key mechanism in the model. Due to saving constraints, farmers have an unmet demand to receive part of the income from their produce after harvest. Competition destroys the relational contract between the farmer and the mill. A key aspect of that relationship is the mill's ability to credibly promise payments after the harvest. Without this promise, farmers process coffee at home in order to save income until after harvest. Note that the lower processing from the mill is *not* picked up by competing mills: the *aggregate* amount of cherries sold by farmers and processed by mills decreases. This is a key distinction between the mechanism in the model and the standard business-stealing mechanism. Under business-stealing competition reduces the volume of coffee processed by each individual mill but increases the aggregate volume of processed coffee in the market.

Column 4 shows that competition increases the likelihood that farmers have to self-finance inputs without increasing overall input usage (unreported). Column 5 shows that competition also does not affect yields, measured as kilos of cherries per tree. Column 6 also shows that competition does not lead farmers to invest in their plantation and increase the number of trees. The evidence thus suggests that competition does not provide farmers with incentives to improve their productivity and/or invest.

The lack of an effect of competition on prices, yields and investment suggests that competition does not increase farmers' returns from coffee cultivation. The effect of competition on farmers profits and welfare is potentially ambiguous. The results shown so far suggest that competition does not increase the revenues farmers receive from coffee cultivation. Competition, however, could decrease the costs incurred by farmers, raise profits and make farmers better off. For example, we have shown above that competition reduces the quality of cherries sold by farmers. Competition might thus save farmers effort and other costs needed to produce quality.

We therefore consider the effect of competition on an overall measure of farmers' welfare. It is notoriously difficult to elicit accurate estimates of profits for farming enterprises in general. These difficulties are particularly pronounced in our context in which farmers' literacy levels are low, accounting records are not kept, the main input on the farm (labour) is difficult to price, and coffee cultivation coexists alongside several other farming and non-farming activities. We therefore focus on an overall index of farmer's satisfaction as our preferred proxy for farmers' welfare. Column 7 in Table 6 shows that competition has a strong negative impact on farmers' overall reported satisfaction.

The evidence in its totality supports the model's predictions on farmer-level outcomes. The evidence also rejects the hypothesis that farmers benefit from competition. If anything, the overall satisfaction score suggests that competition does make the average farmer worse off. This would be consistent with simple extensions of the model in which farmers earn rents under the relational arrangement.[28]

# 6 Mechanisms and Discussion

## 6.1 Mechanisms: "Temptation" vs. "Profits" (Prediction C)

The model highlights *two* distinct mechanisms through which competition erodes mills' ability to sustain relational contracts with farmers: a *direct temptation* mechanism (it is harder to commit to the higher post-harvest transfers needed to induce farmers to induce farmers to sell to the mill rather than being tempted by the competing mill); and an *indirect profit* mechanism (competition reduces the mill's profits and makes it harder to honor the relational contract even with those farmers not directly affected by competition).

We attempt to untangle the two mechanisms in the data. The idea behind the empirical strategy is to distinguish, for each farmer, competition from mills that are near

[28]Unreported results also show that competition reduces trust between farmers and the mill. The survey posed trust questions adapted from the World Value Surveys. Since farmers were interviewed at the mill we could not ask farmers directly about trust in the mill's manager. Following suggestions from our local enumerator team we asked about trust in people from Kigali (the capital city) to capture attitudes towards business people with whom the farmer has a subordinate relationship. We find that competition lowers farmers' trust in mill's management and coffee collectors. Similarly, competition lowers mill's manager trust towards farmers. For farmers and managers alike a placebo test find no effect of competition on general trust, trust in family and trust in neighbors.

the mill *and* the farmer (farmer-competition) and competition from mills that are near the mill but *not* the farmer (mill-competition). Both the direct and indirect effects operate in the first case, while only the indirect effect operates in the second case. The effect of mill-competition thus identifies whether the indirect *profit* mechanism operates. The difference between the effects of farmer-competition and mill-competition identifies whether the direct *temptation* mechanism is at work.

We divide the area surrounding the mill into four quadrants: north-west, north-east, south-east, and south-west. We then assign each farmer to her quadrant. For each farmer, then, we split competition into the number of mills in the farmer's quadrant (farmer-competition) and in the three other quadrants (mill-competition). We instrument competition in the farmer's own quadrant and in the other quadrants with the average suitability score in the relevant regions of the donut.

Table 7 reports the results. Column 1 reports OLS estimates splitting the number of mills within 10 kms from the mill into farmer-competition and mill-competition. The estimates confirm a negative correlation between both measures of competition and the use of relational contracts as reported by the farmer, although neither is precisely estimated. These OLS estimates might be biased upward or downward due to a number of different concerns, including measurement error. Relative to the specification in Table 2, there are two additional sources of measurement error in the specification in Column 1 of Table 7. First, mills in other quadrants might also directly affect the farmer. Second, the process through which farmers are assigned to quadrants is noisy, as we do not have GIS coordinates for the farmer's plot instead we use the centroid coordinates of the farmer's village.[29]

We therefore explore an IV specification in which we separately instrument for farmer-competition and mill-competition using the average suitability score between 5 and 10 km from the mill in the farmer's quadrant and in the three other quadrants as instruments. Columns 2 and 3 in Table 7 report the first stages. Reassuringly, we find that the suitability score in the farmer's quadrant strongly correlates with farmer-competition (p-value $< 0.01$) but not with mill-competition, and vice-versa.

[29]Farmers were surveyed at the mill and thus we only know the name of the village where the farmer's plot is located. This wouldn't per se be a major limitation given that the average village has an area just larger than 1 km$^2$. Unfortunately, however, names do not uniquely identify villages and respondents of different age and ethnicity often refer to the same village using different names. We thus look for each surveyed farmer in the (de-anonymized) version of the national census of coffee farmers to assign farmers to a village and, thus, location. We are able to exactly locate approximately 70% of our surveyed farmers. The table reports results on this sample. We are able to locate an additional 10% of farmers through a fuzzy match procedure and find similar results when including those. The results in the main Tables of the paper are robust if we restrict the analysis to this restricted sample of farmers.

Column 4 reports the IV estimates. We find suggestive evidence that *both* mechanisms are at play. An additional competing mill in the farmer's quadrant reduces the use of relational contracts by nearly 0.3 standard deviations. This is the effect of competition operating through both the direct and indirect effects. An additional mill in other quadrants also reduces the use of relational contracts by 0.19 standard deviations. This is the impact due to the indirect effect only. The difference between the two estimates, nearly 0.1 standard deviations (p-value 0.21), identifies the direct temptation effect.

### 6.2 Discussion of Exclusion Restriction

We conclude with a discussion of threats to, and evidence in support of, our identification strategy. We instrument for competition using the average predicted score from the engineering model in the donut area between 5 and 10 km from the mill. The exclusion restriction is satisfied if, conditional on suitability conditions within the mill's catchment area, average suitability in the 5-10 km area only affects mill's operation through its effect on competition.

A possible concern is that unobservable conditions that directly affect the mill's operation might be correlated with the instrument. For example, farmers' skills and entrepreneurial attitudes, or the presence of well-functioning input and financial markets, or efficient extension services from the government could, in principle, correlate with suitability and with mill's operations. We argue, however, that this is unlikely to be a significant threat to the validity of our identification strategy. First, note that the model includes the suitability score in the catchment area. These unobservables would be of concern if they were correlated with the suitability score outside the catchment area in ways not captured by their correlation with the suitability score inside the catchment area.

Second, these unobservables would need to feature an a priori implausible correlation structure. To see why, consider the reduced form correlation between unit costs and the suitability score at various distances from the mill. As expected, a higher suitability score in the mill's catchment area correlates with *lower* operating costs for the mill. A high suitability score between 5 and 10 km (our instrument), however, correlates with *higher* operating costs for the mill. Beyond 10 km the suitability score has no relationship with the mill's operating costs. Our interpretation is that the *reversal* in the relationship between suitability and operating costs is due to the fact that better conditions between 5 and 10 km attracts more competition. To invalidate our identification structure these unobservable factors would thus also need to be negatively auto-correlated over space and their effect fade at 10 km.

A second concern is that the instrument for competition relies only on cross-sectional variation. Entry of new mills in the sector, however, happens over time. It is therefore important to consider whether dynamic aspects of the entry decisions might invalidate the identification strategy. Given the discussion above, the main threat to our identification strategy is posed by the possibility that certain mills strategically locate in areas with high suitability but surrounded by areas with low suitability ("oases") anticipating lower competition in the future. If those mills are also better at establishing relational contracts with farmers then our identification strategy would not be valid.[30]

Three pieces of evidence and numerous conversations in the field suggest that this threat is unlikely to be of practical relevance. First, we can check how the mills' order of entry correlates with the suitability score within 5 km (catchment area) and between 5 and 10 km (instrument) from the actual mill's entry location. We expect earlier entrants to locate in places with higher score within 5 km from the mill. Under the dynamic strategic entry considerations above we expect earlier entry to locate in areas with worse surrounding. In contrast, the validity of our identification strategy suggests that we should see no correlation between the order of entry and the instrument, conditional on suitability within the catchment area.

Figure A4 lends support to our identification strategy. The Figure confirms that earlier entrants locate in catchment areas with higher suitability while later entrants settle for locations with lower suitability. The corresponding trend for the instrument is also negative. We thus find the opposite of what is implied by strategic dynamic entry considerations. The negative trend in the instrument is driven by spatial correlation: conditional on suitability within the catchment area the instrument is uncorrelated with the order of entry.

Second, we exploit data from the two years following our survey and check whether, conditional on suitability, unit costs of existing mills correlate with the location choice of new entrants. Table A4 reports the results. The unit of observation is a sector, the lowest level for which we know the location of new entrants.[31] The dependent

[30]Note that the cross-sectional nature of our instrument identifies structural conditions that presumably affects mill's ability to sustain relational contracts in the steady state, rather than temporary instability that might be caused by the entry of new competitors. Ghani and Reed (2018) provides a case study describing such instability looking at the evolution of the relationships between fishermen and ice suppliers following the entry of an additional ice manufacturer.

[31]Coincidentally the average sector has an area similar to the baseline definition of catchment area.

variable is a dummy taking value equal one if a new mill has entered that sector in 2013 or 2014 and zero otherwise. We check whether unit costs of existing mills in 2012 predict the location decision of new entrants. Since existing mills' unit costs are observed only where mills existed in 2012, dummies for the existence of mills in the sector in 2012 are included as controls. Column 1 shows that there is no correlation between installed capacity or average unit costs in 2012 and subsequent entry. Column 2 confirms this result adding controls for unit costs and capacity in neighboring sectors. Because Columns 1 and 2 do not control for suitability for entry at the sector level, the coefficient on existing capacity on subsequent entry is likely to be biased upward. Column 3 controls for average suitability score in the sector and finds that, indeed, the suitability score positively predicts entry while installed 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 individual components of the entry model and also confirms the results.

Last, we check the relationship between competition and mill manager's characteristics. This exercise checks whether competition is associated with worse managers and is thus of independent interest, since we might expect competitive pressures to lead mills to select better managers. Table A3 shows that there is no relationship between competition and 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.

In sum, a variety of tests lend empirical support to the validity of our identification strategy. In particular, we find no evidence that mills base their strategic entry decisions on either the suitability for, or the actual performance of, mills in places 5 to 10 km beyond their entry location. Perhaps more importantly, numerous field visits and conversations with investors and regulators give us confidence in our identification strategy. These conversations reveal that indeed investors do consider multiple locations before deciding where to establish a mill. They also report, however, to base their decisions on conditions prevailing in the vicinity of the considered locations. To find potential "oases" (and invalidate our identification strategy) investors would need to scope an area of $(10^2 - 5^2) \times \pi) \approx 235.50\ km^2$ well beyond the locations they are considering. Such extensive scoping would be difficult to undertake in a systematic way. Recall that the information required to assess suitability for mill placement, including geo referenced farmers' censuses and other GIS data, was assembled *ex novo* for this project and was thus not accessible before this study to prospective investors and regulators alike.

## 7 Conclusion

In settings where formal contracting institutions are poor, parties rely on *relational contracts* − informal agreements sustained by the future value of the relationship − to deter short-term opportunism and facilitate trade. Empirical evidence on the scope, structure and determinants of these informal arrangements has the potential to identify the key market failures in specific contexts and inform policy, particularly in a development context.

This paper presents an empirical study of the effect of competition on the relational contracts between coffee mills and farmers in Rwanda, a context that is of intrinsic interest but is also convenient from a methodological point of view. We make two contributions. First, we contribute to the literature on relational contracts and, more broadly, on management practices. We systematically measure relational practices in a sample of large firms; we document significant dispersion in the adoption of these practices; we show these practices are complementary; and confirm that their adoption is strongly correlated with firm's actual performance. Relational practices are, by definition, hard to codify and context-specific. While the practices we measure are relevant in our setting and, more broadly, in agricultural value chains in developing countries, we hope to offer an example of how relational contracts can be measured in other contexts as well.

Second, we study the role of competition as a determinant of the adoption of relational practices. We argue this is the key comparative static to understand whether poor contract enforcement alters market functioning. In a first-best world, we expect competition to have a positive effect on management quality and productivity. A distinctive feature of relational contracts is that rents are relied upon to curb opportunism and, to the extent competition erodes those rents, it could lead to worse outcomes. We find a significant negative impact of competition between mills on the use of relational contracts between mills and farmers. The breakdown in relational contracts lowers mills' efficiency and output quality. More surprisingly, competition between mills lowers the aggregate amount of coffee supplied by farmers to any mill and, if anything, makes farmers worse off. This provides novel evidence on the functioning of markets in second-best environments.

These findings must be interpreted cautiously. Our results demonstrate that in a second-best world the benefits of competition might be hampered by the presence of other market failures which are mitigated by relational contracts. Our analysis identifies the average effect of adding an additional competitor for a mill that is already subject to intense competition. The results should therefore not be interpreted as supporting monopsony.[32]

The evidence suggests the possibility of socially excessive entry when contracts are hard to enforce. A direct policy recommendation from our results, then, is to improve contract enforcement in agricultural chains. While it might be too much of a task to improve a country's formal court system, industry regulators can improve contract enforcement in specific agricultural chains. For example, in Costa Rica the Instituto del Cafe de Costa Rica (ICAFE) is mandated to monitor the coffee value chain and enforces contracts between mills and farmers and between mills and exporters. The conditions for such policy interventions, however, must be evaluated on a case-by-case basis. First, improvements in contract enforcement can lead to the first-best only if they are sufficiently large. Partial improvements in contract enforcement could actually worsen market functioning by further undermining relationships (Baker et al. (1994)). Second, contract enforcement will alter the distribution of rents across actors in the chain. Political economy aspects must thus be taken into account in the design of such reforms (Paige (1997)). In light of our results, further research is needed to understand how to effectively improve contract enforcement in these contexts.

[32]The findings could, potentially, offer a rationale for regulations commonly observed in agricultural chains in developing countries, e.g., zoning and minimum distance rules. These are of course prone to abuse too.

## References


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

[2] Andrabi, T., Das, J., and Khwaja, I. (2017) "Report Cards: The Impact of Providing School and Child Test Scores on Educational Markets." *American Economic Review*, 107(6), 1535-63.

[3] Antras, P. (2015) *Global Production: Firms, Contracts, and Trade Structure.* Princeton University Press.

[4] Antras, P., and Foley, C. (2015) "Poultry in Motion: A Study of International Trade Finance Practices." *Journal of Political Economy*, 123(4), 809-852.

[5] Arrow, K. (1962) "Economic Welfare and the Allocation of Resources for Invention," in Richard R. Nelson, ed., *The Rate and Direction of Inventive Activity.* Princeton, NJ: Princeton University Press, pp. 609-26.

[6] Baker, G., Gibbons, R., and Murphy, K.J. (1994) "Subjective Performance Measures in Optimal Incentive Contract." *Quarterly Journal of Economics*, 109(4), 1125-1156.

[7] Baker, G., Gibbons, R., and Murphy, K. J. (2002) "Relational Contracts and the Theory of the Firm." *Quarterly Journal of Economics*, 117(1), 39-84

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

[9] Bardhan, P. K. (Ed.) (1989) *The Economic Theory of Agrarian Institutions*. Oxford: Clarendon Press.

[10] Binswanger, H. P., and Rosenzweig, M. R. (1986) "Behavioural and Material Determinants of Production Relations in Agriculture." *Journal of Development Studies*, 22(3), 503-539.

[11] Bloom, N., and Van Reenen, J. (2007) "Measuring and Explaining Management Practices across Firms and Countries", *Journal of Economic Perspectives*, 24(1),Winter 2010.

[12] Bloom, N., and Van Reenen, J. (2010) "Why do Management Practices Differ across Firms and Countries." *Annual Review of Economics* 2(1), 105-137.

[13] Bloom, N., Propper, C., Seiler, S., and Van Reenen, J. (2015) "The Impact of Competition on Management Practices in Public Hospitals." *Review of Economic Studies*, 82 (2), 457-489.

[14] Bloom, N., Brynjolfsson, E., Foster, L., Jarmin, R., Patnaik, M., Saporta-Eksten, I., and Van Reenen, J. (2017) "What Drives Differences in Management?", NBER Working Paper No. 23300.

[15] Blouin, A., and Macchiavello, R. (forthcoming) "Strategic Default in the International Coffee Market." *Quarterly Journal of Economics*

[16] Casaburi, L., and Reed, T. (2017) "Competition in Agricultural markets: An Experimental Approach", Working Paper, Zurich

[17] Casaburi, L., and Macchiavello, R. (2019) "Demand and Supply of Infrequent Payments as Commitment Device: Evidence from the Kenyan Dairy Industry." *American Economic Review*, 109 (2), 523-555

[18] Casaburi, L., and Willis, J. (2018) "Time versus State in Insurance: Experimental Evidence from Contract Farming in Kenya." *American Economic Review*, 108 (12), 3778-3813.

[19] Chetverikov, D., Kim, D., and Wilhelm, D. (forthcoming) "Nonparametric instrumental variable estimation." *Stata Journal*

[20] Conley, T. (1999) "GMM Estimation with Cross Sectional Dependence," *Journal of Econometrics*, 92 (1), 1-45.

[21] Djankov, S., La Porta, R., Lopez-de-Silanes, F., and Schleifer, A. (2003) "Courts." *Quarterly Journal of Economics*, 118 (2), 453-517.

[22] Ghani, T., and Reed, T. (2018) "Relationships on the Rocks: Contract Evolution in a Market for Ice," Working Paper, Washington University Olin Business School.

[23] Gibbons, R., and Henderson, R. (2012) "What Do Managers Do?: Exploring Persistent Performance Differences Among Seemingly Similar Enterprises". In *The Handbook of Organizational Economics*, edited by Gibbons, R., and Roberts, J., 680-731. Princeton University Press.

[24] Gil, R., and Zanarone, G. (2015) “Testing for the Interaction of Formal and Informal Contracts”, Working Paper, Queen’s University

[25] Ghosh, P., and Ray, D. (1996) “Cooperation in Community Interaction without Information flows.” *Review of Economic Studies*, 63 (3), 491-519.

[26] Greif, A. (1993) “Contract Enforceability and Economic Institutions in the Early Trade: The Maghribi traders’ coalition.” *American Economic Review*, 83 (3), 525-548.

[27] Hall, P., and Horowitz, J. (2005) “Nonparametric Methods for Inference in the Presence of Instrumental Variables.” *The Annals of Statistics*, 33 (6), 2904-2929.

[28] Helper, S., and Henderson, R. (2014) “Management Practices, Relational Contracts, and the Decline of General Motors.” *Journal of Economic Perspectives*, 28 (1), 49-72.

[29] Holmes, T. J., and Schmitz Jr, J. A. (2010) “Competition and Productivity: A Review of Evidence.” *The Annual Review of Economics*, 2 (1), 619-642.

[30] Hsieh, C., and Olken, B. (2014) “The Missing ’Missing Middle”, *Journal of Economic Perspective*, 28(3), 89-108

[31] Jensen, R. T., and Miller N. H. (2018) “Market Integration, Demand, and the Growth of Firms: Evidence from a Natural Experiment in India.” *American Economic Review*, 108 (12), 3583-3625.

[32] Konig, M.D., Rohner, D., Thoenig, M., and Zilibotti, F. (2017) “Networks in Conflict: Theory and Evidence From the Great War of Africa.” *Econometrica*, 85 (4), 1093?1132.

[33] Kranton, R. E. (1996) “Reciprocal Exchange: a Self-Sustaining system.” *American Economic Review*, 86 (4), 830-851.

[34] Levin, J. (2003) “Relational incentive contracts.” *American Economic Review*, 93 (3), 835-857.

[35] Macchiavello, R., and Morjaria, A. (2015) “The Value of Relationships: Evidence from a Supply Shock to Kenyan Rose Exports.” *American Economic Review*, 105 (9), 2911-45.

[36] Macchiavello, R., and Morjaria, A. (2015) "Fully Washed Coffee Exports in Rwanda: Market Structure and Policy Implications", International Growth Centre, Policy Brief.

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

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

[39] Paige, J. M. (1997). *Coffee and Power: Revolution and the Rise of Democracy in Central America*. Cambridge, Mass.: Harvard University Press.

[40] Petersen, M. A., and Rajan, R. G. (1995) "The Effect of Credit Market Competition on Lending Relationships." *Quarterly Journal of Economics*, 407-443.

[41] Rodrik, D. (2008) "Second-Best Institutions." *American Economic Review*, 98 (2), 100-104.

[42] Schilling, M., and McConnell. T. (2006) "Model for Siting Coffee Washing in Rwanda, ESRI Paper 1697

[43] Schmitz Jr., J. A. (2005) "What Determines Productivity? Lessons from the Dramatic Recovery of the U.S. and Canadian Iron Ore Industries Following Their Early 1980s Crisis." *Journal of Political Economy*, 113 (3), 582-625.

[44] Schumpeter, J. (1942) *Capitalism, Socialism and Democracy.* New York: Harper and Brothers

[45] Söderbom, M., and Kamarudeen, S. (2013) "Constraints and opportunities in Rwanda's Industrial Sector." Working Paper, International Growth Centre.

[46] Startz, M. (2018) "The Value of Face-to-Face: Search and Contracting Problems in Nigerian trade." Working Paper, Stanford

[47] Stigler, G. (1956) "The Statistics of Monopoly and Merger." *Journal of Political Economy*, 64 (1), 33-40.

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

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

# Appendix

## 7.1 Data

The empirical analysis combines an original census of mills, conducted via a comprehensive survey, with administrative and geographical data. There are three main sets of data. First, we assembled a high resolution (at the 1 $km^2$ cell) GIS database with information on geographic, climatic and infrastructure characteristics for the whole of Rwanda. This is essential to construct environmental control variables and estimate the engineering model for mill placement. Second, we matched the Rwanda coffee census conducted in 2008/9 with the GIS data. The census covers all farmers coffee trees and their village location.[33] This census provides basic information about trees and production at a highly disaggregated level for places with and without mills. Additional details on the survey and the other sources of data are provided below.

### 7.1.1 Mill Census 2012

The survey was designed by the authors in collaboration with the National Agricultural Exporting Board and was implemented by one of the authors towards the end of the 2012 harvest season, between May and July 2012. The survey covered all operating mills in the 2012 harvest season. The response rate was nearly 100% (due to heavy rain and poor road conditions one mill could not be surveyed). Four survey teams each including six team members were trained by one of the authors to administer the survey and cover the whole of Rwanda over the season. Each survey team was led by a qualified coffee personnel from the coffee board. Upon previous appointment, face-to-face interviews were conducted. The average survey completion time was 4 hours per mill and each team was able to survey two mills a day.

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 characteristics 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 including financial and geographic details. Finally, at each mill a representative sample of the mill's output, the coffee parchment was obtained to be physically and chemically analyzed in the coffee board's quality laboratory in Kigali.

The manager (and owner if necessary) survey module included for general questions

[33] The average village is a bit larger than 1 $km^2$. This allows us a precise match of the coffee census to the GIS data.

about the mill (separate modules for private or cooperative mills, mill history, assets owned by the mill, NGO assistance, operational activity, procurement of cherries, relationship with farmers, employment, access to finance, marketing and capital invested) and the manager itself (socio-economic characteristics, canonical World Value Survey questions, employment history, earnings from station and income profile, incentive payments, raven and numerical test).[34]

Five random farmers were selected to be interviewed at each mill. The farmers were extracted from the coffee board's district officer list of farmers in the sector. Farmers were asked about their socio-economic characteristics, raven and numerical tests were conducted, various questions related to their coffee production (ownership of trees, past and current production and sales), membership in cooperatives, interlinked transactions, input and canonical trust questions.

Four random workers were also interviewed at the mill. Workers were randomly chosen on the spot by the survey team on arrival at the mill. The main coffee collector for each mill was also surveyed. These modules are not used for this paper.

Sample of coffee lots were also taken from each mill to be physically examined and cupped in the NAEB laboratory in Kigali. Survey team leaders choose one random lot at the mill. The lot was then assigned a random code before being taken to the lab so that the mill could not be identified by the cuppers at the laboratory. The coding was done by one of the authors.

Lastly a GPS module of the mill survey collected information on the mills location and elevation.

#### 7.1.2 GIS Data

Several ancillary data sources were obtained from coffee board as well as other agencies (Ministry of Agriculture, Rwanda Natural Resources Authority, Ministry of Infrastructure, Rwanda Meteorological Agency, National University of Rwanda). GIS tools allows us to divide Rwanda into a one square kilometer cell (Rwanda's size is approx. 25,000 $km^2$). We are able to assign each square kilometer, its geographic characteristics (altitude, slope, historical coffee suitability, different types of roads and rivers) by extracting data from using remote sensing technologies on ortho-photos of the country.

[34]During the visit we obtained photographic copies for most of mill's their records, including attendance records, wage bills, payments for cherry deliveries as well as sales and loan contracts. Summary from these documents allowed a careful checking of unit costs figures. We have not digitalized the photographic records. We have used them to check some of the responses from the survey. When the mill's manager was only in charge of production and sourcing, we followed up with interviews with owners in Kigali to elicit information about marketing and financial aspects.

Using a detailed Geographic Information System (GIS) dataset extracted from orthophotos taken in 2008/09 (aerial pictures at the 0.25m pixel size) we are able to map out all the rivers of different hierarchies in Rwanda, in particular our interest is on the small rivers which form the end points of springs. From this master river database we first isolate the smallest rivers and obtain their origins by triangulating information on elevation using the Shutter Radar Topography Mission raster data which allows us at 10 m resolutions to ascertain the elevation and hence end points of springs. We then overlay this on a GIS map of Rwanda that has divided the country into 1 $km^2$ boxes (refereed to as cell's in the main text). This allows us to determine whether a cell has a spring and also allows us to determine the distance to the closet spring.

### 7.1.3 Note on Measurement of Unit Costs

Coffee mills are characterized by a relatively simple technology. It takes approximately 5.50 kgs of coffee cherries to produce 1 kg of parchment coffee. The exact conversion ratio depends on coffee variety and other geographical factors affecting the organic properties of coffee. We 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 mills accounts at the end of the season together with the mill managers, we obtained accurate measures of unit costs and their breakdown across components. Assuming a Leontieff production function, unit costs of mill $i$ can be written as follows:

$$UC_i = \left(P_i^{kg} \times CR_i\right) + OC_i \tag{10}$$

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. Labour, transport and procurement costs are relatively easy to compute from the accounts. As usual, capital costs require additional assumptions. We cross-check the figures reported by the manager with the (marginal) interest rate paid by the mill on working capital loans. As expected, there is almost 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 in unit costs originates from the components that are more directly influenced by management: labour, capital, procurement and logistic. Here we find a 90/10 ratio equal to 2.32.

## 7.2 Proofs

***Proof of Observation 3***

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 (4), we obtain $P_2 = \rho\,(q_r - q/2) + e$.

If any relational contract can be sustained at all, it must then entail $q_m = q$. For suppose that there exists a $\widetilde{q}_r \in (q/2, q)$ such that constraint (5) 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. The assumption can of course be relaxed at the cost of keeping track of an additional case with an interior solution without gaining any further insight ||.

Table 1: DESCRIPTIVE STATISTICS

| | (1) Mean | (2) Median | (3) $25^{th}$ Pct. | (4) $75^{th}$ Pct. | (5) Obs |
|---|---|---|---|---|---|
| | - | - | - | - | - |
| *Panel A: Mill Characteristics* | | | | | |
| Theoretical Capacity (tons of cherries) | 428.652 | 500 | 250 | 500 | 178 |
| Production (tons of green coffee) | 46.015 | 32 | 15 | 60 | 177 |
| Cherries purchased (tons) | 294.818 | 199.910 | 102.380 | 400 | 174 |
| Seasonal employees | 35.135 | 30 | 16 | 50 | 171 |
| Farmers in catchment area that sell to mill | 395.959 | 310 | 170 | 500 | 170 |
| Cooperative status | 0.466 | 0 | 0 | 1 | 178 |
| Age of mill | 4.090 | 4 | 2 | 6 | 178 |
| Unit cost (RWF per kg) | 1792.989 | 1800 | 1600 | 1956 | 178 |
| Unit cost processing (RWF per kg) | 705.340 | 699 | 500 | 831 | 177 |
| Number of mills within 10km | 6.348 | 6 | 3 | 9 | 178 |
| *Given inputs to farmers* | 0.222 | 0 | 0 | 0 | 176 |
| *Has made a second payment in the past* | 0.784 | 1 | 1 | 1 | 176 |
| *Provides help/loans to farmers* | 0.773 | 1 | 1 | 1 | 176 |
| Relational contracting z-score | 0 | 0.114 | -0.502 | 0.453 | 175 |
| Engineering z-score within 5km | 0 | -0.424 | -0.787 | 0.887 | 177 |
| Engineering z-score 5-10km | 0 | -0.220 | -0.854 | 0.821 | 177 |
| Average elevation (m) within 5km | 1624.452 | 1626.961 | 1510.328 | 1729.064 | 177 |
| Average slope within 5km | 10.933 | 10.872 | 8.835 | 12.910 | 177 |
| Average river density within 5km | 320.664 | 322.843 | 197.796 | 424.854 | 177 |
| Average tree density within 5km | 11831.191 | 9507.520 | 5141.521 | 14762.032 | 177 |
| Average spring presence within 5km | 0.034 | 0.029 | 0.0 | 0.058 | 177 |
| Meters of road within 5km | 1771.712 | 1685.932 | 1431.913 | 1967.669 | 177 |
| Coffee quality index, overall | 0 | 0.129 | -0.408 | 0.766 | 140 |
| *Panel B: Farmer Characteristics* | | | | | |
| Age | 46.445 | 47 | 36 | 56 | 875 |
| Female | 0.287 | 0 | 0 | 1 | 881 |
| Years of schooling | 5.339 | 6 | 4 | 7 | 879 |
| Distance to mill, km | 5.480 | 2.689 | 1.194 | 7.182 | 615 |
| Cooperative membership | 0.552 | 1 | 0 | 1 | 881 |
| Trees owned | 975.518 | 500 | 250 | 1000 | 881 |
| *Received input from mill* | 0.176 | 0 | 0 | 0 | 881 |
| *Expects to receive a second payment* | 0.795 | 1 | 1 | 1 | 881 |
| *Expects to receive help/loan* | 0.637 | 1 | 0 | 1 | 877 |
| Cherry price (RWF per kg) | 201.057 | 200 | 190 | 220 | 878 |
| Job Satisfaction index | 0 | 0.026 | -0.457 | 0.499 | 868 |
| Number of other mills in own quadrant | 1.784 | 2 | 0 | 3 | 519 |

*Note:* Mill characteristics are obtained from the census of mill survey. Farmer characteristics are obtained from 4-5 random farmers supplying to the surveyed mill. Both surveys took place at the same time and were fielded in the harvest season of 2012; Text in *italics* refers to the relational contract measures. See Data Appendix for additional details on the survey.

Table 2: COMPETITION AND RELATIONAL CONTRACTS: IV ESTIMATES

| | (1)<br>RC Score<br>(z-score) | (2)<br>Competition | (3)<br>RC Score<br>(z-score) | (4)<br>RC Score<br>(z-score) |
|---|---|---|---|---|
| Competition | -0.112<br>(0.027)***<br><0.026>***<br>[0.028]***<br>[0.028]*** | | | -0.254<br>(0.065)***<br><0.058>***<br>[0.051]***<br>{0.043}*** |
| Score<br>within 5-10 km of mill | | 2.072<br>(0.277)***<br><0.279>***<br>[0.291]*** | -0.527<br>(0.132)***<br><0.130>***<br>[0.121]*** | |
| Score within 5 km of mill | YES | YES | YES | YES |
| Geographic controls | YES | YES | YES | YES |
| Mill controls | YES | YES | YES | YES |
| adjusted $R^2$ | .28 | .76 | .29 | .19 |
| N | 175 | 175 | 175 | 175 |

*Note:* Standard errors are denoted as follows – (Bootstrap in which mills are resampled with replacement and the regression is repeated to generate the distribution of the coefficient); <Standard errors adjusted for arbitrary spatial clustering using the *acreg* package written by Konig and coauthors and used in Konig et al. (2017)>; [Standard errors that adjust for spatial clustering as in Conley (1999), implemented by Conley's *x_gmm Stata* package]; {Subsampling bootstrap, where we randomly subsample 90% of cell points in the index construction stage to reestimate the engineering model and scores, and then rerun the second stage regression. We resample 1000 times, and report the standard deviation of the estimates.} $***$ ($**$) [$*$] indicates significance at the 0.01 (0.05) [0.1] level. All regressions control for NGO-supported, cooperative status, station age and station age squared. Mill controls also include average engineering score, average spring presence, road density, tree density, rivers, coffee suitability, elevation, slope within 5 kilometers, and latitude and longitude coordinates. Pre-harvest z-score is constructed based on farmer- and mill manager- based indicators of mill-provided inputs. Harvest z-score is constructed from farmer- and mill manager- based indicators of second payments post-harvest. Post-harvest z-score is constructed from farmer- and mill manager- based indicators loans or help provided after the harvest. The RC Score is an aggregate of these three indexes. Competition is measured as the number of mills within 10 km, and is instrumented with the engineering model score in locations 5km to 10km away from the mill. See Data appendix for additional details on the survey and GIS datasets.

Table 3: UNPACKING COMPLEMENTARY RELATIONAL PRACTICES

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) | (13) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Received input from mill | Given inputs to farmers | Pre-Harvest Z-score | Expects to receive a second payment | Has made a second payment in the past | Harvest Z-score | Expects to receive help/loan | Provides help/loans to farmers | Post-Harvest Z-score | RC Score | RC Score, farmer outcomes | RC Score, mill outcomes | Placebo: Short-Term Credit |
| **Panel A: IV** | | | | | | | | | | | | | |
| Competition | -0.045*** | -0.067** | -0.171** | -0.062*** | -0.067*** | -0.190** | -0.060*** | -0.041 | -0.191** | -0.254*** | -0.204*** | -0.199*** | -0.011 |
| | (0.011) | (0.027) | (0.067) | (0.013) | (0.026) | (0.080) | (0.016) | (0.034) | (0.081) | (0.072) | (0.033) | (0.066) | (0.083) |
| **Panel B: OLS** | | | | | | | | | | | | | |
| Competition | -0.010* | -0.032** | -0.062** | -0.032*** | -0.037*** | -0.106*** | -0.022*** | -0.021 | -0.077*** | -0.112*** | -0.078*** | -0.104*** | -0.049 |
| | (0.005) | (0.013) | (0.027) | (0.007) | (0.013) | (0.034) | (0.008) | (0.016) | (0.028) | (0.030) | (0.014) | (0.032) | (0.039) |
| Score within 5 km of mill | YES | 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 | YES |
| Mill controls | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Farmer controls | YES | – | – | YES | – | – | YES | – | – | – | YES | – | – |
| Adjusted $R^2$ | 0.12 | 0.21 | 0.21 | 0.16 | 0.19 | 0.24 | 0.05 | -0.04 | 0.03 | 0.19 | 0.14 | 0.13 | 0.07 |
| N | 869 | 176 | 176 | 869 | 176 | 176 | 865 | 175 | 175 | 175 | 869 | 176 | 172 |

*Note:* Bootstrapped standard errors in parentheses. $***$ ($**$) [$*$] indicates significance at the 0.01 (0.05) [0.1] level. Mill controls include NGO-supported, cooperative status, station age and station age squared. Geographic control includes average engineering score, average spring presence, road density, tree density, rivers, coffee suitability, elevation, slope within 5 kilometers, and latitude and longitude coordinates. Pre-harvest z-score is constructed based on farmer- and mill manager- based indicators of mill-provided inputs. Harvest z-score is constructed from farmer- and mill manager- based indicators of second payments post-harvest. Post-harvest z-score is constructed from farmer- and mill manager- based indicators loans or help provided after the harvest. The RC Score is an aggregate of these three indexes. Farmer controls include farmer age, education, gender, schooling, distance to mill, cognitive z-score, cooperative membership. Farmer responses all come from the farmer survey. Competition is measured as the number of mills within 10 km, and is instrumented with the engineering model score in locations 5km to 10km away from the mill.

Table 4: COMPETITION AND MILL OUTCOMES

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
|---|---|---|---|---|---|---|---|---|---|---|
| | Unit Costs (ln) | Cherry Price (ln) | Conversion Ratio (ln) | Processing Unit Costs (ln) | Capacity Utiliza- tion | Weeks Mill Pro- cessed | Days w/out enough cherries | Unit Labour Costs (ln) | Days with too many workers | Days with too few workers |
| **Panel A: IV** | | | | | | | | | | |
| Competition | 0.037*** | -0.001 | -0.009 | 0.072** | -4.653 | -0.106 | 0.056 | 0.149*** | 0.331*** | 0.150** |
| | (0.014) | (0.009) | (0.012) | (0.029) | (3.187) | (0.460) | (0.035) | (0.037) | (0.097) | (0.062) |
| **Panel B: OLS** | | | | | | | | | | |
| Competition | 0.010* | 0.002 | -0.011** | 0.016 | -0.595 | 0.059 | 0.028* | 0.066*** | 0.099** | 0.031 |
| | (0.005) | (0.004) | (0.005) | (0.010) | (1.505) | (0.169) | (0.014) | (0.021) | (0.043) | (0.025) |
| Score within 5 km of mill | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Engineering controls | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Geographic controls | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Mill controls | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Adjusted $R^2$ | -0.04 | 0.28 | 0.00 | -0.06 | -0.01 | 0.10 | 0.13 | 0.11 | 0.00 | -0.07 |
| N | 177 | 163 | 145 | 176 | 169 | 160 | 169 | 176 | 173 | 174 |

*Note:* Bootstrapped standard errors in parentheses. $***$ ($**$) [$*$] indicates significance at the 0.01 (0.05) [0.1] level. Mill controls include NGO-supported, cooperative status, station age and station age squared. Engineering controls and Geographical controls include average engineering score, average spring presence, road density, tree density, rivers, coffee suitability, elevation, slope within 5 kilometers, and latitude and longitude coordinates. "Days without enough cherries", "Days with too many workers" and "Days with too few workers" are all dummy variables from manager survey. Competition is measured as the number of mills within 10 km, and is instrumented with the engineering model score in locations 5km to 10km away from the mill. All responses come from the mill manager survey.

Table 5: COMPETITION AND COFFEE QUALITY

| | (1) Overall Quality Score | (2) Farmer-Controlled Quality | (3) Mill-Controlled Quality | (4) Plant Genetic Quality |
|---|---|---|---|---|
| **Panel A: IV** | | | | |
| Competition | -0.149* | -0.171** | -0.107 | 0.029 |
| | (0.081) | (0.075) | (0.086) | (0.060) |
| **Panel B: OLS** | | | | |
| Competition | -0.066* | -0.051 | -0.054 | 0.015 |
| | (0.037) | (0.034) | (0.046) | (0.025) |
| Score within 5 km of mill | YES | YES | YES | YES |
| Engineering controls | YES | YES | YES | YES |
| Geographic controls | YES | YES | YES | YES |
| Mill controls | YES | YES | YES | YES |
| Adjusted $R^2$ | 0.00 | 0.04 | 0.13 | -0.03 |
| N | 140 | 155 | 156 | 157 |

*Note:* Bootstrapped standard errors in parentheses. $***$ ($**$) [$*$] indicates significance at the 0.01 (0.05) [0.1] level. Mill controls include NGO-supported, cooperative status, station age and station age squared. Engineering controls and Geographical controls include average engineering score, average spring presence, road density, tree density, rivers, coffee suitability, elevation, slope within 5 kilometers, and latitude and longitude coordinates. Competition is measured as the number of mills within 10 km, and is instrumented with the engineering model score in locations 5km to 10km away from the mill. The farmer-controlled quality index is a standardized index of an indicator of large beans, and of severe insect damage. The mill-controlled quality index is constructed from an indicator of high moisture, of floaters, and of broken beans. The overall quality score is constructed from the farmer and mill indices, plus an indicator of ideal conversion ratio, an indicator of specialty status, and standardized cupping points. All components of indices are rescaled so that higher values indicate higher quality. See Data Appendix for additional details on coffee lot sampling.

Table 6: COMPETITION AND FARMER OUTCOMES

| | (1) Cherry Price (ln) | (2) Share sold as cherries | (3) Home process for saving | (4) Self financed inputs | (5) Yield (ln) | (6) Farmer's trees (ln) | (7) Job Satisfaction Index |
|---|---|---|---|---|---|---|---|
| **Panel A: IV** | | | | | | | |
| Competition | 0.008* | -0.015** | 0.051*** | 0.074*** | 0.005 | 0.008 | -0.083*** |
| | (0.005) | (0.007) | (0.015) | (0.017) | (0.026) | (0.041) | (0.020) |
| **Panel B: OLS** | | | | | | | |
| Competition | 0.003 | -0.008** | 0.012*** | 0.026*** | 0.035*** | 0.000 | -0.032*** |
| | (0.003) | (0.004) | (0.007) | (0.008) | (0.011) | (0.000) | (0.007) |
| Score within 5 km of mill | YES | YES | YES | YES | YES | YES | YES |
| Engineering controls | YES | YES | YES | YES | YES | YES | YES |
| Geographic controls | YES | YES | YES | YES | YES | YES | YES |
| Mill controls | YES | YES | YES | YES | YES | YES | YES |
| Farmer controls | YES | YES | YES | YES | YES | YES | YES |
| Adjusted $R^2$ | 0.10 | 0.07 | 0.04 | 0.15 | 0.15 | 0.04 | 0.09 |
| N | 866 | 842 | 869 | 865 | 869 | 869 | 856 |

*Note:* Bootstrapped standard errors in parentheses. $***$ ($**$) [$*$] indicates significance at the 0.01 (0.05) [0.1] level. Mill controls include NGO-supported, cooperative status, station age and station age squared. Engineering controls and Geographical controls include average engineering score, average spring presence, road density, tree density, rivers, coffee suitability, elevation, slope within 5 kilometers, and latitude and longitude coordinates. Competition is measured as the number of mills within 10 km, and is instrumented with the engineering model score in locations 5km to 10km away from the mill. Farmer controls include farmer age, education, gender, schooling, distance to mill, cognitive score, cooperative membership, and log number of trees (except in column 5). "Job Satisfaction Index is a standardized index of 1-4 (4 being strongly agree) responses of agreements to statements: "My job gives me a chance to do the things I do best", "The pay is good", "I learn new things", "I am treated with respect", "I have freedom to do my work", and "I do not find work stressful". All farmer outcomes come from the farmer survey.

Table 7: RELATIONAL CONTRACTING AND COMPETITION: FARMER LEVEL

| | (1) RC score, farmer outcomes | (2) Farmer Competition | (3) Mill Competition | (4) RC score, farmer outcomes |
|---|---|---|---|---|
| Farmer-level Competition | -0.041 | | | -0.294** |
| | (0.031) | | | (0.114) |
| Mill-level Competition | -0.107*** | | | -0.191*** |
| | (0.025) | | | (0.073) |
| Suitability score in own quadrant | | 0.864*** | -0.256 | |
| | | (0.150) | (0.188) | |
| Suitability score in other quadrants | | -0.193 | 1.896*** | |
| | | (0.141) | (0.405) | |
| Equality of Coefficients, p-value | | | | 0.216 |
| Competition Measure | # mills in quadrant (farmer), # mills in other quadrants (mill) | - | - | # mills in quadrant (farmer), # mills in other quadrants (mill) |
| Instrument | - | - | - | Score in quadrant, score in other quadrants |
| Mill Controls | YES | YES | YES | YES |
| District FE | YES | YES | YES | YES |
| Farmer Controls | YES | YES | YES | YES |
| Quadrant Fixed Effects | YES | YES | YES | YES |
| OLS/IV | OLS | OLS | OLS | IV |
| Adjusted $R^2$ | 0.24 | 0.38 | 0.62 | 0.10 |
| N | 511 | 511 | 511 | 511 |

*Note:* Standard errors clustered at mill level. $***$ ($**$) [$*$] indicates significance at the 0.01 (0.05) [0.1] level. Farmer controls include distance to mill, farmer age, education, gender, schooling, cognitive score, cooperative membership, and log number of trees. Mill controls include NGO-supported, cooperative status, station age and station age squared, as well as average engineering score, average spring presence, road density, tree density, rivers, coffee suitability, elevation, slope within 5 kilometers, and latitude and longitude coordinates. Quadrants are adjacent 10 km by 10 km squares with the mill situated at the shared corner. Quadrant fixed effects refer to dummies for the Northeast, Southeast, etc. quadrant. Main farmer outcomes included in the RC score are receiving fertilizer from the mill, receiving a second payment, and expecting a loan or help after harvest. The number of observations falls when quadrant-level competition is introduced, because some farmers do not reside in any quadrant (they are more than 10 kilometers from the mill). P-value is reported for the test that farmer and mill effects are equal.

## Table A1: ENGINEERING MODEL

| | (1) Mill | (2) Mill | (3) Mill | (4) Mill |
|---|---|---|---|---|
| Spring within cell | 0.358** | | 0.359** | -3.084*** |
| | (0.170) | | (0.171) | (0.251) |
| Untarred Local Road within cell | | 0.411*** | 0.412*** | 0.389*** |
| | | (0.143) | (0.143) | (0.143) |
| Interaction | | | | 3.467*** |
| | | | | (0.261) |
| **Geographic Controls:** | | | | |
| Polynomials | YES | YES | YES | YES |
| Interactions | YES | YES | YES | YES |
| pseudo-$R^2$ | 0.12 | 0.12 | 0.12 | 0.12 |
| N | 13970 | 13970 | 13970 | 13970 |

*Note:* Standard errors are clustered by sector. $***$ ($**$) [$*$] indicates significance at the 0.01 (0.05) [0.1] level. Observations are cells on the map of Rwanda. All regressions control for linear, quadratic, and cubic terms of elevation, slope, coffee suitability, rivers, the number of trees in sector, and latitude and longitudinal coordinates, as well as the interactions between each of these variables (indicated by "Interaction"). See Data Appendix for GIS data construction.

## Table A2: COMPETITION AND ACCESS TO CREDIT

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | Any working capital loan | Expected interest rate on 100 mil. | Loans from buyers | Extreme difficulty borrowing 100 mil | Share of 100 million to purchase cherries | Return on 100 mil exceeds interest |
| **Panel A: IV** | | | | | | |
| Competition | 0.084* | 0.226 | 0.052 | 0.133** | -0.005 | 0.044 |
| | (0.047) | (0.221) | (0.034) | (0.052) | (0.019) | (0.029) |
| **Panel B: OLS** | | | | | | |
| Competition | 0.018 | -0.039 | 0.011 | 0.039* | -0.007 | 0.010 |
| | (0.018) | (0.101) | (0.014) | (0.020) | (0.008) | (0.010) |
| Score within 5 km of mill | YES | YES | YES | YES | YES | YES |
| Engineering controls | YES | YES | YES | YES | YES | YES |
| Geographic controls | YES | YES | YES | YES | YES | YES |
| Mill controls | YES | YES | YES | YES | YES | YES |
| Adjusted $R^2$ | . | 0.04 | 0.12 | -0.05 | 0.02 | . |
| N | 177 | 176 | 174 | 173 | 167 | 171 |

Bootstrapped standard errors in parentheses. $***$ ($**$) [$*$] indicates significance at the 0.01 (0.05) [0.1] level. Mill controls include NGO-supported, cooperative status, station age and station age squared. Engineering controls and Geographical controls include average engineering score, average spring presence, road density, tree density, rivers, coffee suitability, elevation, slope within 5 kilometers, and latitude and longitude coordinates. Competition is measured as the number of mills within 10 km, and is instrumented with the engineering model score in locations 5km to 10km away from the mill. "Any working capital loan" and "Loan from buyers" are dummy variables. All responses come from the manager survey.

Table A3: COMPETITION AND MANAGEMENT CHARACTERISTICS

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---|---|---|---|---|---|---|---|---|
| | Age | Education | Cog Tests | Tenure at Mill | Months Full-time | Received Training | Income | Incentive |
| **Panel A: IV** | | | | | | | | |
| Competition | -0.565 | 0.017 | -0.077 | 0.110 | -0.110 | -0.021 | -0.019 | -0.016 |
| | (0.905) | (0.089) | (0.055) | (0.136) | (0.300) | (0.035) | (0.055) | (0.026) |
| **Panel B: OLS** | | | | | | | | |
| Competition | -0.446 | 0.017 | -0.013 | -0.024 | 0.034 | -0.001 | 0.003 | -0.007 |
| | (0.387) | (0.038) | (0.031) | (0.054) | (0.120) | (0.013) | (0.029) | (0.011) |
| Score within 5 km of mill | YES | YES | YES | YES | YES | YES | YES | YES |
| Engineering controls | YES | YES | YES | YES | YES | YES | YES | YES |
| Geographic controls | YES | YES | YES | YES | YES | YES | YES | YES |
| Mill controls | YES | YES | YES | YES | YES | YES | YES | YES |
| Adjusted $R^2$ | 0.03 | 0.05 | -0.02 | 0.16 | 0.23 | 0.07 | 0.46 | 0.05 |
| N | 167 | 177 | 177 | 174 | 174 | 177 | 171 | 177 |

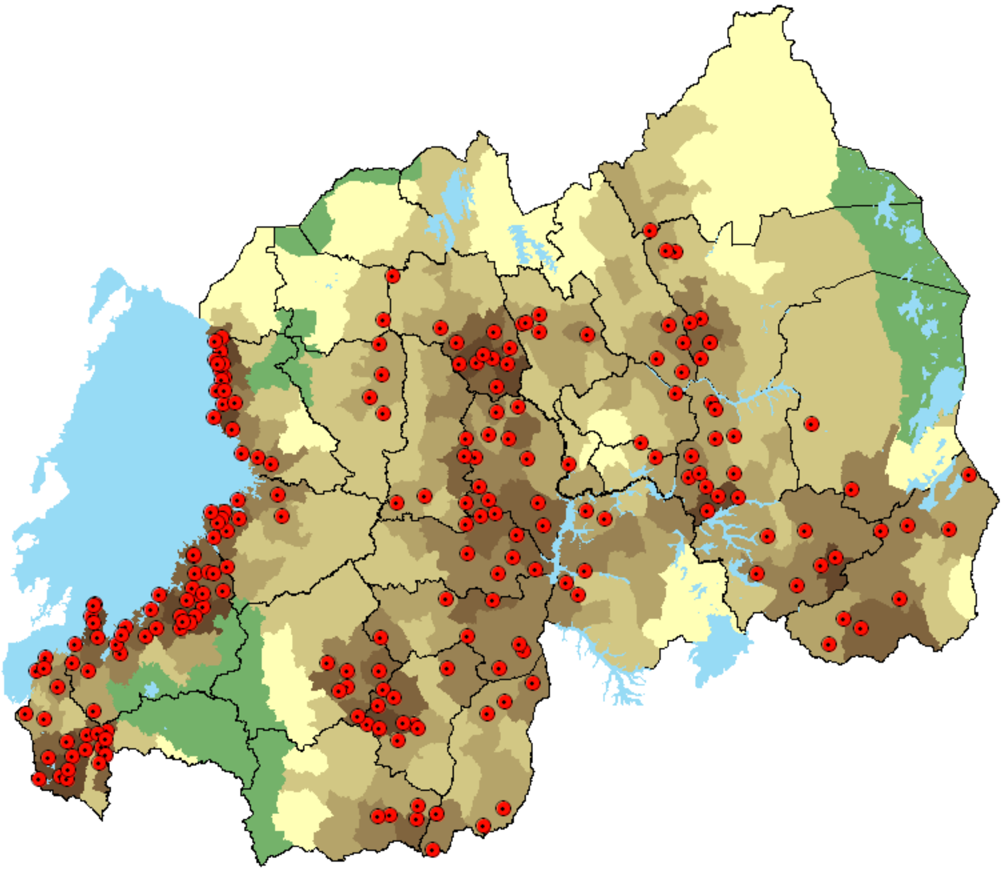
Bootstrapped standard errors in parentheses. $***$ $(**)$ $[*]$ indicates significance at the 0.01 (0.05) [0.1] level. Mill controls include NGO-supported, cooperative status, station age and station age squared. Engineering controls and Geographical controls include average engineering score, average spring presence, road density, tree density, rivers, coffee suitability, elevation, slope within 5 kilometers, and latitude and longitude coordinates. Competition is measured as the number of mills within 10 km, and is instrumented with the engineering model score in locations 5km to 10km away from the mill. All responses come from the manager survey.

Table A4: MILL UNIT COST AND SUBSEQUENT ENTRY

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| **Dep. Var.: Entry of New Mills in the Years 2013 and 2014** | | | | |
| Unit Costs of Mill in Sector, 2012 | -0.006 | -0.024 | 0.024 | 0.050 |
| | (0.172) | (0.170) | (0.168) | (0.158) |
| Capacity in Sector, 2012 | 59.863 | 54.622 | 4.590 | -6.857 |
| | (42.425) | (44.462) | (58.691) | (55.146) |
| Unit Costs of Mill in Neighbouring Sectors, 2012 | | 0.080 | 0.068 | 0.079 |
| | | (0.057) | (0.054) | (0.059) |
| Capacity in Neighbouring Sectors, 2012 | | 3.071 | -15.746* | -16.723* |
| | | (7.881) | (8.362) | (9.098) |
| Average Score in Sector | | | 5.487*** | 3.436* |
| | | | (1.606) | (1.806) |
| Geographic and Engineering Controls: | NO | NO | NO | YES |
| adjusted $R^2$ | 0.07 | 0.07 | 0.12 | 0.12 |
| N | 416 | 416 | 413 | 413 |

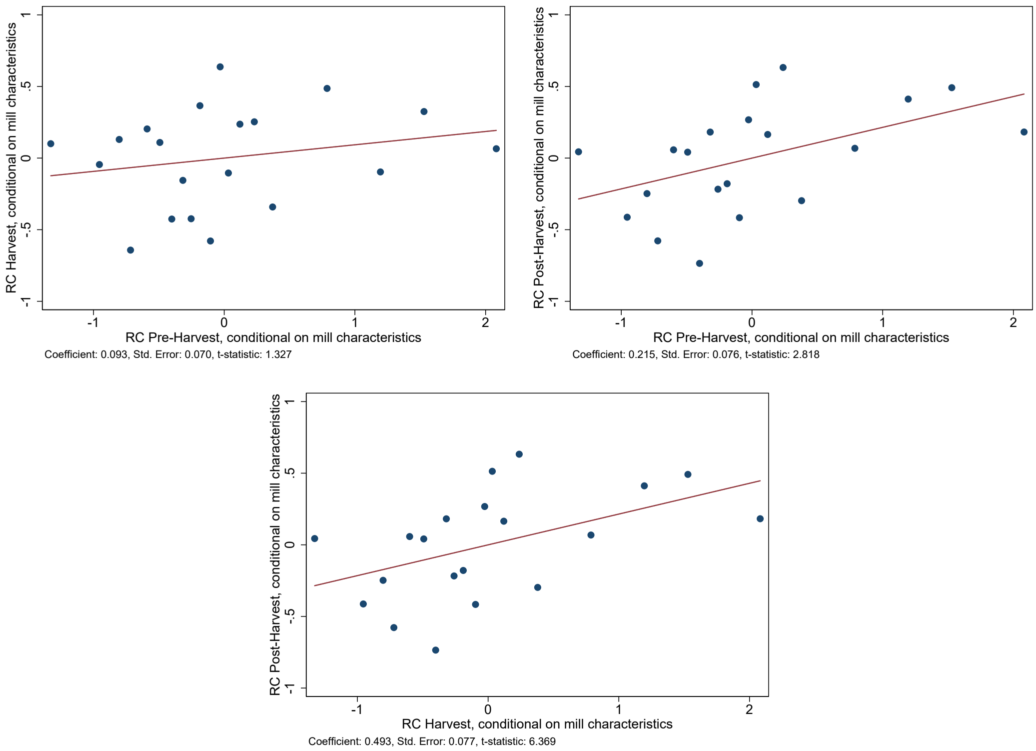
*Note:* Standard errors are clustered by district. $***$ ($**$) [$*$] indicates significance at the 0.01 (0.05) [0.1] level. Observations are sectors of Rwanda. Sectors are the third-level administrative units in Rwanda, with an area of approximately 50 $km^2$. Geographic and engineering controls are tree density, sector size, road density, spring present, elevation, and slope.

Figure 1: MILL PLACEMENT IN RWANDA, 2012



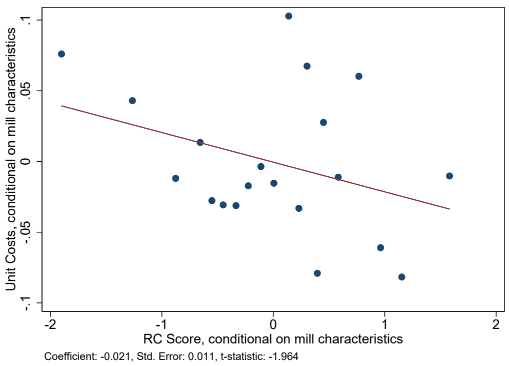
*Note:* This figure illustrates the spatial distribution of mills in Rwanda in 2012 (denoted by red dots). In the 2012 harvest season there were in total 211 mills in Rwanda. The green shade indicates national parks and blue indicates water bodies. The background overlay is the number of coffee trees at the sector level (the third administrative unit of Rwanda) and the darker the shade of brown the higher the number of coffee trees in the sector. This figure is for illustration purposes only. Source: author's survey of mils and various other data sources, see Data Appendix for additional details.

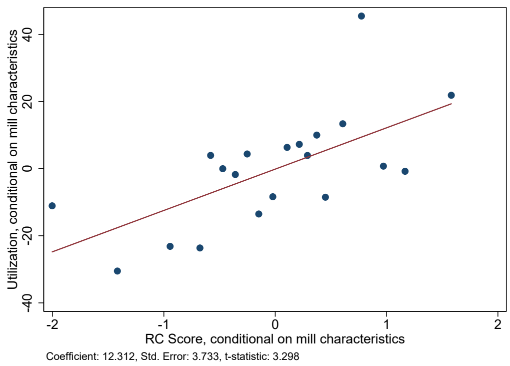
Figure 2: CORRELATIONS BETWEEN RELATIONAL CONTRACT PRACTICES




*Note:* Binned scatter plot of mill-level regressions. All regressions control for NGO-support, cooperative status, station age and station age squared. Controls also include average engineering score, average spring presence, road density, tree density, rivers, coffee suitability, elevation, slope within 5 kilometers, and latitude and longitude coordinates. Pre-harvest z-score is constructed based on farmer- and mill manager- based indicators of mill-provided inputs. Harvest z-score is constructed from farmer- and mill manager- based indicators of trade credit and second payments. Post-harvest z-score is constructed from farmer- and mill manager- based indicators of loans and/or help provided to farmers unrelated to harvest operations. Source: author's survey of mills and various other data sources, see Data Appendix for additional details.

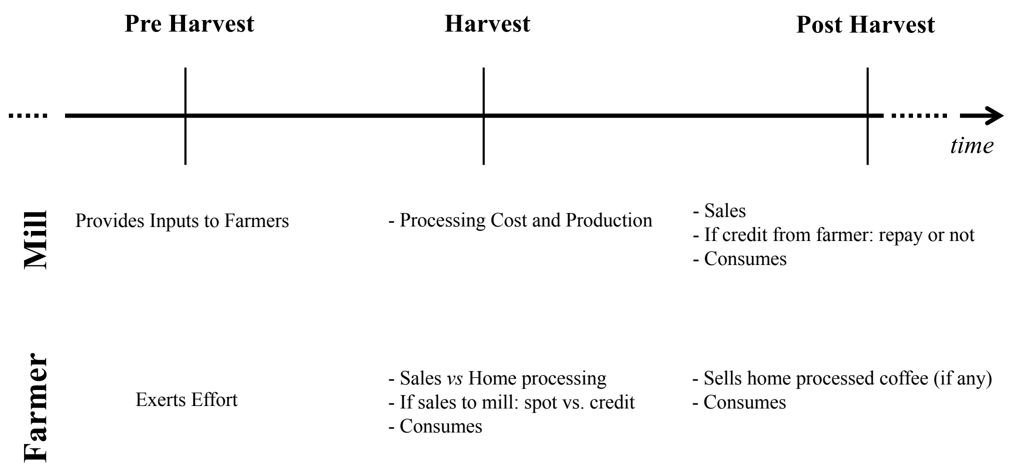
Figure 3: VALIDATING RELATIONAL CONTRACT SCORE: UNIT COST OF PROCESSING AND UTILIZATION OF MILL



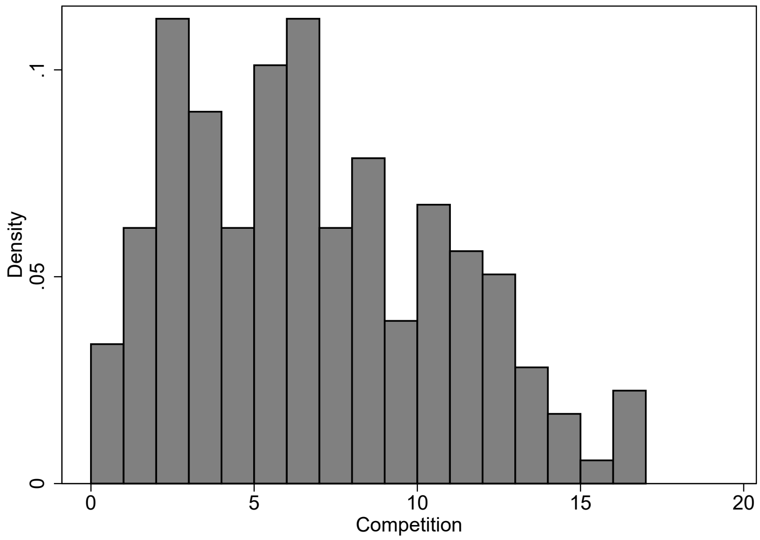

*Note:* Binned scatter plot of mill-level regressions. All regressions control for NGO-support, cooperative status, station age and station age squared. Controls also include average engineering score, average spring presence, road density, tree density, rivers, coffee suitability, elevation, slope within 5 kilometers, and latitude and longitude coordinates. The RC Score is an aggregate of farmer- and mill manager- based indicators of mill-provided inputs, of second payments, and of post-harvest loans. Unit costs are operating costs (in Rwandese Francs) per Kg of parchment produced. Capacity utilization is tons of cherries processed in the season divided by theoretical capacity. Source: author's survey of mils and various other data sources, see Data Appendix for additional details.

Figure 4: TIMING OF EVENTS DURING A COFFEE SEASON



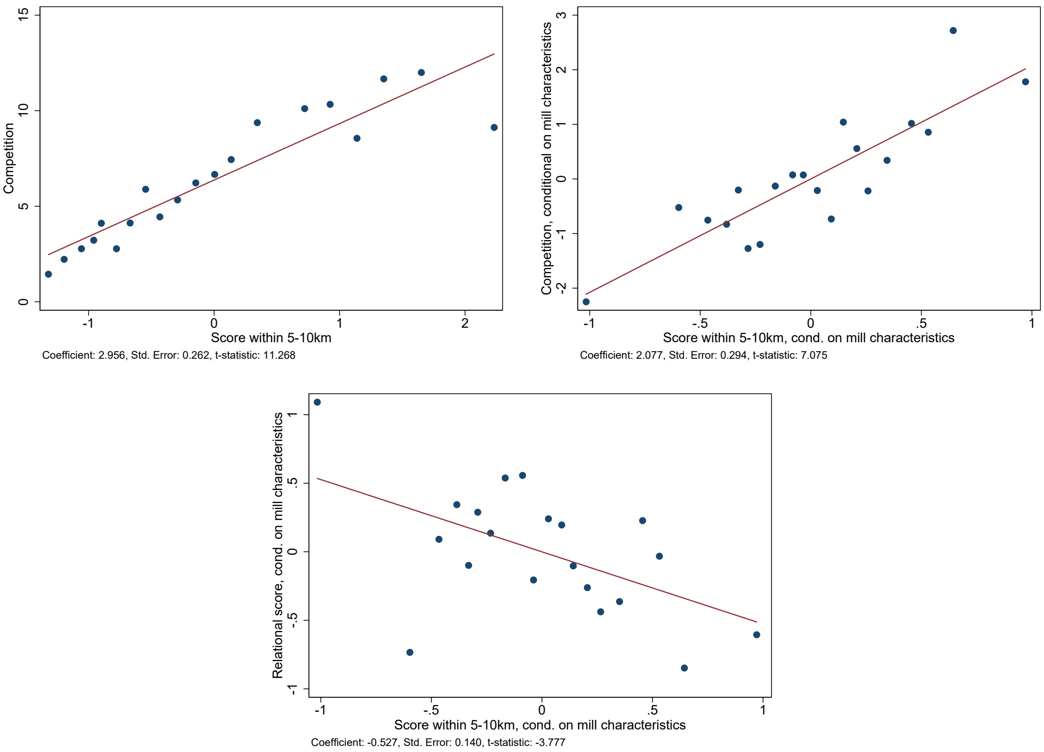

*Note:* The Figure depicts the timing of events in the model. Time is an infinite sequence of identical seasons. Each season is divided into three sub-periods: pre-harvest, harvest, and post-harvest. Prior to harvest, the mill decides whether to provide inputs to farmers or not. Farmers then decide whether to exert effort. At harvest, production is realized. The farmer decides whether to sell to the mill or to home process the coffee. If the farmer home processes the coffee, she decides how much to sell for current consumption and how much to store until post-harvest. If the farmer sells any coffee to the mill, the mill processes it and, together with the farmer, agrees the timing of payments. Finally, post-harvest, the mill sells the coffee and decides to make any payment to farmer or default. The farmer consumes her income: payments from the mills (if any) and/or sales from stored home processed coffee (if any).

Figure 5: COMPETITION BETWEEN MILLS WITHIN 10KM RADIUS



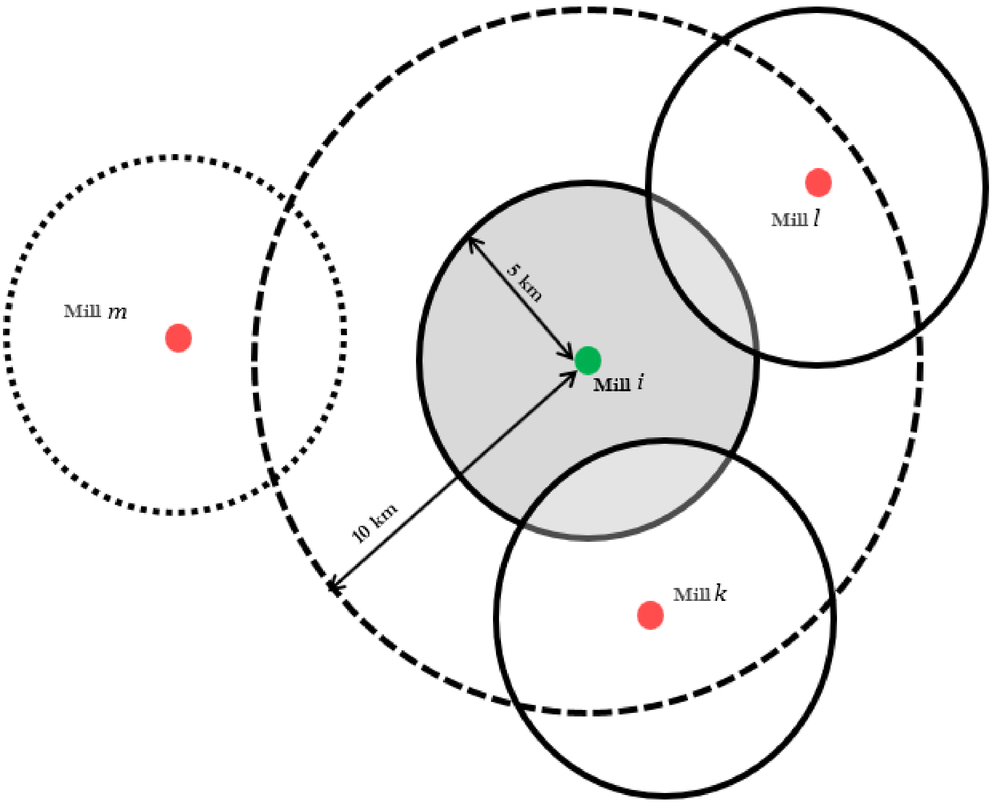

*Note:* The figure plots the distribution of the number of competing mills within 10 km of each mill. Source: author's calculation from geo-coded coordinates of mills and 2012 mill survey.

Figure 6: PARTIAL REGRESSION PLOTS: FIRST STAGE AND REDUCED FORM



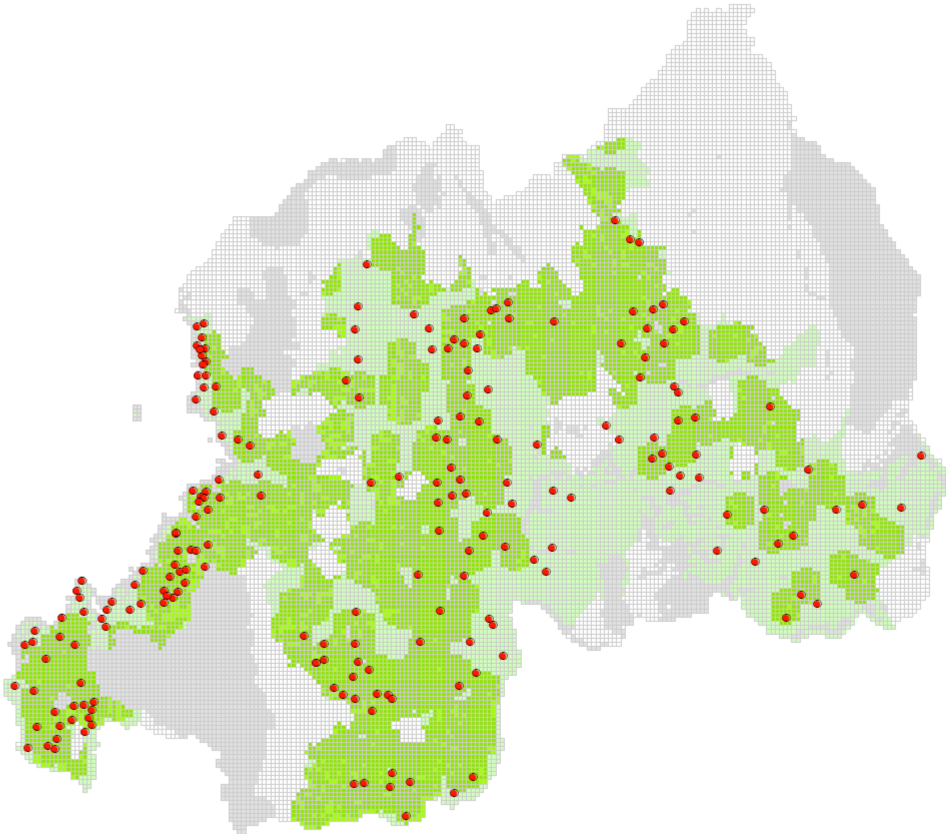

*Note:* Binned scatter plot of mill-level regressions. All regressions control for NGO-support, cooperative status, station age and station age squared. Controls also include average engineering score, average spring presence, road density, tree density, rivers, coffee suitability, elevation, slope within 5 kilometers, and latitude and longitude coordinates. The RC Score is an aggregate of farmer- and mill manager- based indicators of mill-provided inputs, second payments, and post-harvest loans. Competition is measured as the number of mills within 10 km.

Figure A1: GRAPHICAL REPRESENTATION OF COMPETITION MEASURE



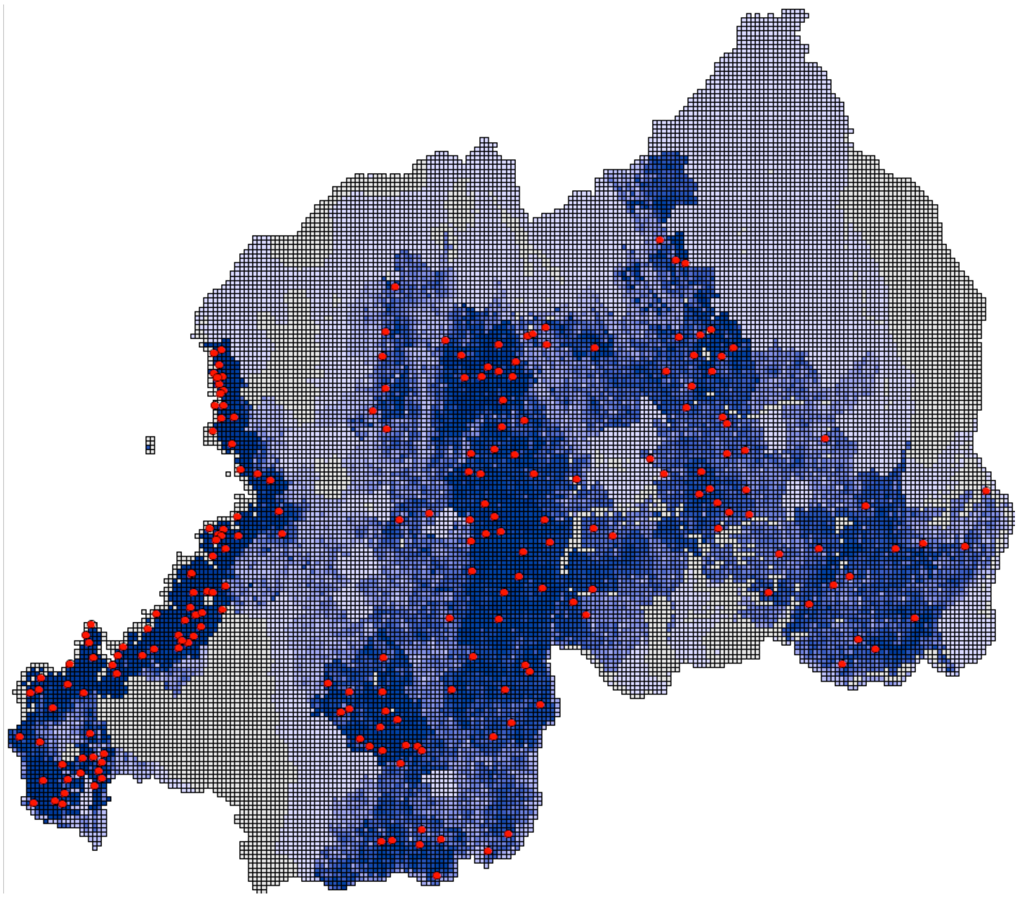

*Note:* This figure illustrates the 5km catchment area for mill $i$. Any mill within a 10km radius of mill $i$ will have a catchment area that overlaps (at least to some extent) with mill $i$'s catchment area. The overlap is illustrated in the graph for mill $k$ and $l$. Our competition measure based on a 5km catchment area therefore includes all mills within a 10km radius. This is represented by the dashed circle in the figure.

Figure A2(A): ENGINEERING MODEL: CRITERIA AND MILLS



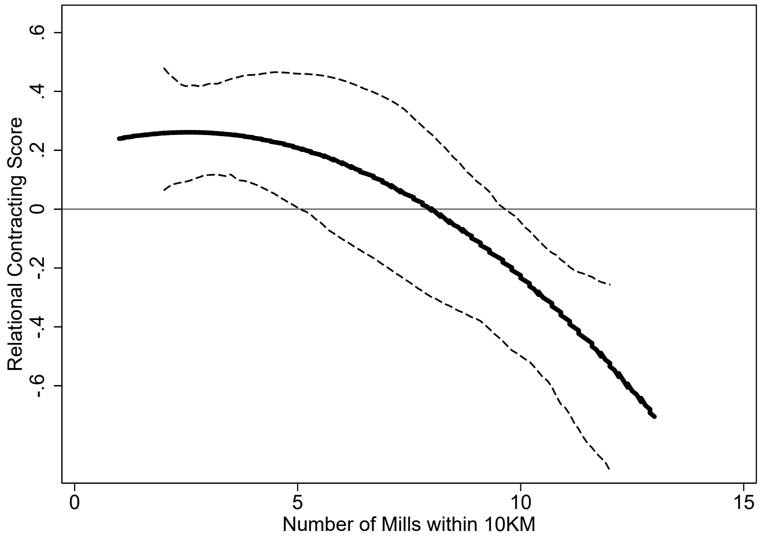
*Note:* This figure illustrates the engineering model's criteria: the dark grey cells are ineligible for mill placement due to presence of national parks, water body or are built-up areas. The lightest green illustrates cells that satisfy the number of trees necessary for mill placement, the brightest green areas highlight where the cells 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 Data Appendix for additional details.

Figure A2(B): ENGINEERING MODEL: MILL PLACEMENT



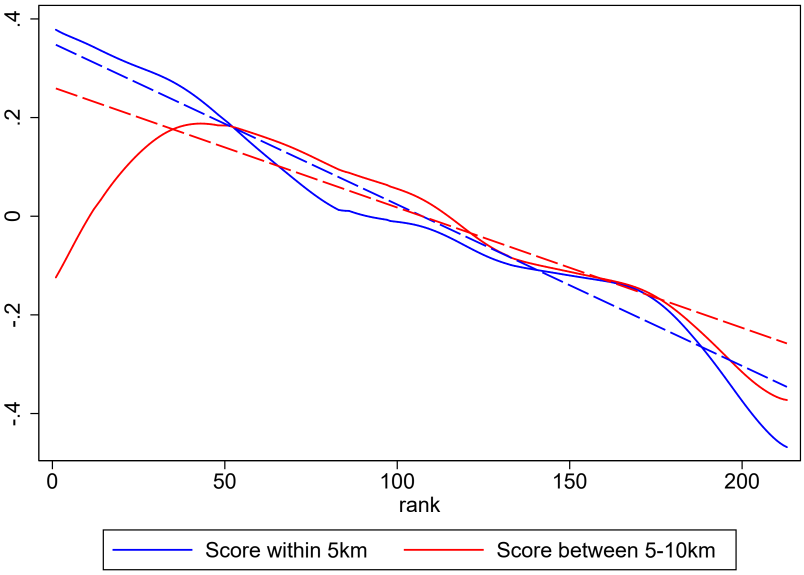
*Note:* This figure illustrates, the predicted "score" for the placement of a mill in each cell (1 $km^2$) in Rwanda using our model of mill placement, which is driven by engineering considerations for the optimal placement of mills. The darker the color higher the probability of mill placement. Red dots illustrate existing mills. Source: author's calculation on various GIS datasets, see Data Appendix for additional details.

Figure A3: NON-PARAMETRIC IV




*Note:* 95% confidence interval, represented by dotted-lines, is based on bootstrap resampling. Non-parametric IV follows Hall and Horowitz (2005) as implemented by *npivreg* authored by Chetverikov et al. (forthcoming). Briefly: (a) control variables are partialled out of the outcome, the endogenous regressor, and the instrument; (b) then polynomial bases for the endogenous regressor and the instrument are computed, and the usual 2-stage least squares problem is solved using these bases; and (b) finally, we compute fitted values over a fine cell of the endogenous regressor values. The non-parametric IV regression controls for NGO-supported status, cooperative status, station age and station age squared. Controls also include average engineering score, average spring presence, road density, tree density, rivers, coffee suitability, elevation, slope within 5 kilometers, and latitude and longitude coordinates. The RC Score is an aggregate of farmer- and mill manager- based indicators of mill-provided inputs, of second payments, and of post-harvest loans.

Figure A4: MILL ORDER OF ENTRY AND SCORE




*Note:* The figure plots a lowess (solid) and linear-fit (dashed) of average score within the catchment area ($< 5$ km) and around it (between 5 and 10 km) against the order of entry. The figure shows that earlier entrants located in better areas (higher average $<5$ km score) but do not appear to have chosen location according to average score between 5 and 10 km. Regressions results confirm that, once controlling for score within 5 km, score between 5 and 10 km does not correlate with the order of entry.