# Competition and Relational Contracts in the Rwanda Coffee Chain*

Rocco Macchiavello† Ameet Morjaria‡

November 2020

**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 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' placement 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. An additional competing mill also (iv) reduces the aggregate quantity of coffee supplied to mills by farmers and (v) makes farmers worse off. Competition hampers relational contracts directly by increasing farmers' temptation to default on the relational contract and indirectly by reducing mill's profits.

*This project was made possible by the collaboration of the Rwanda National Agricultural Exporting Board. Without implicating them, we are especially grateful to Alex Kanyankole, Celestin Gatarayiha and Eric Rukwaya. We owe special thanks to the editors, Pol Antras and Larry Katz and four anonymous referees for many comments and suggestions. Oriana Bandiera, Tim Besley, Robin Burgess, Bob Gibbons, Gerard Padró i Miquel and John Sutton provided invaluable advice and support. We also thank Hunt Allcott, David Atkin, Dan Barron, Nick Bloom, Emily Breza (discussant), Lorenzo Casaburi, Àureo de Paula (discussant), William Fuchs (discussant), Chang-Tai Hsieh (discussant), Asim Khwaja, Adrien Matray (discussant), Dilip Mookherjee, Andy Newman, Mike Powell, Andrea Pozzi (discussant), Steve Tadelis, and Chris Woodruff and participants at seminars in Berkeley-Haas, Bristol, Boston U, Boston College, Cambridge, CEMFI, Chicago Harris, Columbia, Cornell, Georgetown, Gothenburg, HBS, Harvard, Kellogg MEDS, 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 2014 (LSE), CEPR Entrepreneurship 2015, CSIO-IDEI NU-Toulouse 2015, CEPR-WB GVC 2016, CIREQ Applied Economic Conference 2017, DIME-CEGA 2015 (Kigali), EWESM 2014 (Madrid), FOM 2015 (Chicago Booth), IGC Growth Conference 2015, ISNIE 2014 (Duke), NBER Org Meeting (Stanford), NEUDC 2014 (BU), SED 2015 (Warsaw), Stanford CASBS, SSDEV 2014 (Italy) and $2^{nd}$ Workshop on Relational Contracts (Madrid). Financial support from the IGC and the Dean's Research Award at Kellogg School of Management 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 tempting parties with alternative trading opportunities and reducing profits, it may weaken relational contracting and reduce efficiency. What is the impact of competition in such second-best institutional environments?

Answering this question empirically has been challenging for two reasons: first, relational contracts are implicit and context-specific, making such contracts difficult to measure; second, identification of the causal effects of competition is complicated by the endogeneity of market structure. This article identifies the effect of increased competition on firm outcomes in a weakly institutionalized environment in which relational contracts are needed to sustain trade. We address 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, due to poorly functioning input and financial markets typical of agriculture in developing countries' (see, e.g., Bardhan (1989)), sourcing of coffee cherries from farmers at harvest is bundled with legally unenforceable provision of services before, during and after harvest. We measure the use of these relational contracts by conducting an original survey of both mills and farmers in the sector.

Second, we construct an instrument for competition, building on 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 we argue affects relational contracting only through the intensity of mill competition.

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

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; (ii) use fewer relational contracts with farmers; and (iii) exhibit worse performance. 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. We find that competition hampers relational contracts directly by tempting farmers to default on the relational contract and indirectly by reducing mill's profits.

These findings must be interpreted cautiously. We identify the effect of an additional competitor for a mill that competes with six 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".[2] The finding that increased competition downstream leaves all market participants – 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 efficiency.

The article proceeds as follows. Section 2 provides industry background and presents our measure of relational contracts between mills and farmers. In our context a relational contract is a legally non-binding agreement between a mill and supplying farmers that describes how farmers and mills should behave over the course of the entire coffee season. 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 (that is, post) harvest. We measure the use of each relational practice surveying mills' managers and randomly sampled farmers. We aggregate the mill's manager and farmers' responses into a relational contracts index. The relational contracts index displays significant variation and correlates well with mills' performance, giving us confidence that it measures relevant practices for this industry.

Section 3 presents a theoretical framework that captures the key aspects of the relationship between mills and farmers. The model isolates two distinct channels. First, there is a *direct* effect through which competition between mills increases farmers' temptation to renege on the relational contract. Second, competition reduces mills processed volumes and profits. This makes it harder to sustain relational contracts with farmers, even those for which the temptation to renege has not increased. We label this the *indirect* effect. Higher competition might reduce parties' ability to sustain

[2] Adam Smith, (1776), *The Wealth of Nations*, Book I, Chapter XI.

a relational contract. When this occurs, the model delivers a cluster of additional predictions, including an aggregate reduction in cherries procured by mills and lower welfare for 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 the mechanisms.

With regard to the role of competition in sustaining relational contracting, we begin by outlining stylized features of coffee production. Coffee cherries must be processed within hours of harvest and roads are often in poor conditions. Mills thus mainly compete with nearby mills. We measure competition as the number of mills within a 10 km radius from the mill and find that competition negatively correlates with the relational contract index. Ordinary least square (OLS) estimates, however, are likely biased: 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 could itself be 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 and farmers' 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 of engineers and agronomists developed a model to identify suitable sites for mill 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 to locations satisfying the model's criteria. We assembled ex-novo the data required for the model. We predict actual mill placement with the model's criteria and other controls obtaining a "suitability score" for a mill's placement at the 1 $km^2$ resolution for the whole of Rwanda. For each mill we aggregate the suitability score in the area within 5 km radius (henceforth, the catchment area) and in the surrounding area between 5 and 10 km from the mill (henceforth, the instrument). The exclusion restriction is satisfied if, conditional on suitability within the mill's catchment area, suitability in the surrounding area affects mill and farmers' operation only through its impact on competition.

The instrument yields a strong first stage and the second stage finds that an additional mill within 10 km reduces the relational contract index at the mill by 0.28 standard deviations, suggesting that competition has a negative impact on relational contracts. We discuss next extensive evidence mitigating concerns about violations of the exclusion restriction. First, presence of roads and local density of coffee trees are among the variables used to predict the suitability score at the 1 $km^2$. These variables could potentially violate the exclusion restriction and be bad controls. We show that we can omit road density or coffee tree density or both from the construction of the instrument without altering the results. We can also omit those variables from the set of controls in the mill's catchment area without affecting our estimates. Second, our instrument could correlate with farmers' economic opportunities outside coffee, thus reducing the demand for relational contracts. We show that the instrument is uncorrelated with farmers' outside economic opportunities, including access to agricultural markets, to labor markets opportunities, and to financial services.[3]

The evidence in Section 4 suggests that competition leads to a breakdown in relational contracts between mills and farmers. When this happens the model yields a cluster of additional predictions. Section 5 tests these predictions and finds ample support. First, competition reduces relational practices before, during and after harvest by a nearly identical magnitude. Second, at the mill level, competition lowers the amount of cherries processed by the mill and leads to more irregular procurement of cherries. This results in higher average processing cost. We also detect a negative impact of competition on lab-tested quality of random samples of coffee produced by mills, particularly on quality dimensions that depend on farmers' practices. Third, at the farmer level, competition lowers the amount of coffee that farmers sell to any mill without increasing output nor prices. When farmers do not sell cherries to mills, they home process. Given the lower prices fetched by home-processed coffee, an additional mill reduces farmers revenues by about 8%.

Finally, Section 6 investigates mechanisms. We provide evidence consistent with both mechanisms highlighted in the model being at work. Conditional on the number of competing mills and on the farmer's distance to the mill, the relational contract index is lower when the farmer is closer to competing mills. This is consistent with competition directly affecting the farmer's temptation to renege on the relational contract. We also show that, conditional on the number of competing mills to which the farmer can sell to, higher competition from mills to which the farmer cannot sell also reduces the

[3]The Online Appendix B and D explores robustness to alternative definitions of competition, the size of mills' catchment areas, to alternative assumptions on the structure of the error term and to additional threats to identification, including strategic entry effects and differences in market access.

relational contract index. This is consistent with competition indirectly affecting the relational contract through its negative impact on mill's performance.[4] Concluding remarks and policy implications are discussed in Section 7.

This article contributes to three strands of literature. First, to the literature on relational contracts and, more broadly, on management practices (Bloom and Van Reenen (2007, 2010)). The work by Bloom and Van Reenen (2007, 2010) shows that the adoption of certain well-codified management practices is associated with better firm's performance. This 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 practices depends on relational contracts both within and across the firm's boundaries (Baker et al. (2002), Gibbons and Henderson (2012), Helper and Henderson (2014)). Relational practices are, by definition, hard to codify, context-specific and, therefore, hard to measure. This article provides an example of how relational practices can be systematically measured; documents significant dispersion in the adoption of complementary relational practices among firms competing in a narrowly defined industry; and confirms that their adoption correlates with firm performance.[5]

Second, we study the effect of competition in an environment characterized by poor contract enforcement.[6] There is abundant evidence that competition is associated with higher productivity and better management practices. For example, Syverson (2004) shows that in the US larger, more competitive markets are associated with stronger selection in concrete manufacturing. Schmitz Jr. (2005) shows that, in response to competition from Brazilian producers, U.S. iron ore manufacturers increased efficiency and adjusted working arrangements (see also Bloom et al. (2015, 2017) on competition and better management practices). In developing countries, Andrabi et al. (2017), Jensen and Miller (2018) show positive effects of competition on schools in Pakistan and boat builders in Kerala, India respectively. These papers study institutionally

[4]The finding that an additional competitor reduces the aggregate volume of coffee supplied by farmers to mills distinguishes our mechanism from Mankiw and Whinston (1986). In Mankiw and Whinston (1986) an additional entrant increases the total quantity of the good sold in the market and makes consumers better off. Adapting the logic to our context, these predictions are inconsistent with our findings that competition reduces the aggregate quantity processed by mills and makes farmers worse off. Online Appendix D provides additional evidence on this point.

[5]A growing literature studies relationships between firms, often in the context of international markets (see, e.g., Banerjee and Duflo (2000), Macchiavello (2010), Antras and Foley (2015), Macchiavello and Morjaria (2015b), Macchiavello and Miquel-Florensa (2019), Blouin and Macchiavello (2019), Startz (2018), Cajal et al. (2020)). This literature highlights how relationships mitigate contracting problems due to lack of enforcement and/or asymmetric information. We complement this agenda asking how competition affects the sustainability of these relationships.

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

developed environments or contexts in which relational contracts are not key. Our analysis suggests that the benefits of competition might be hampered by the presence of other market failures which are mitigated by relational contracts.

Third, the article relates to the literature on how competition affects relational lending and trade credit. Petersen and Rajan (1995) is a seminal article on how competition might be detrimental to relational lending. McMillan and Woodruff (1999) provides empirical evidence on how firms' outside options affect the ability to sustain relational agreements in a context characterized by weak contract enforcement. Fisman and Raturi (2004) find that monopoly power is negatively associated with credit provision, using data on supply relationships in five African countries. Our article differs from these contributions in several ways. First, we instrument for smoother changes in competition within an oligopolistic setting. Second, we study a context with two sided moral-hazard: both mills and farmers can cheat, just at different points during the harvest season. In contrast, the trade credit literature often considers one-sided moral hazard (suppliers offering trade credit to buyers) and thus when competition increases firms might compete extending trade credit. Ghani and Reed (2020) find evidence consistent with this mechanism exploiting the sudden entry of a new ice manufacturer in Sierra Leone. Casaburi and Reed (2020) find that traders that were randomly offered higher resale prices extended more credit to farmers.[7]

# 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 Rwanda's industry. At the time of our survey in 2012 there were around 350,000 smallholder farmers growing coffee, coffee accounted for almost 20% of the country's exports and between 12% to 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

[7] There has been renewed interest in interlinked transactions in agricultural chains in developing countries (see, e.g, Emran et al. (2020) Casaburi and Macchiavello (2019), Casaburi and Willis (2018) for recent contributions). The literature typically focuses on a single interlinkage at a time (credit, saving, insurance) while we focus on bundles of complementary interlinked transactions and study how they are affected by competition.

harvested. The harvest season typically lasts three to four months and its timing varies across regions depending on altitude and rainfall patterns. Coffee cherries are 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.

Upon harvest, 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, farmers clean cherries at home using rocks before drying then on mats. This process produces coffee cherries of lower and less consistent quality. By contrast, cherries processed through the wet method are taken to a mill (often referred to as coffee washing stations or wet mills) within hours of harvest. If not taken immediately, the cherries will 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. Those who are further afield bring cherries to collection sites in which coffee collectors buy coffee.

The wet method requires specific equipment and substantial quantities of clean 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. 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.[8]

The wet method yields significantly higher value addition for the Rwandan coffee chain as a whole. At the time of our survey, export gate prices for wet-processed coffee (known as fully washed coffee) were around 40% higher than for dry-processed coffee (see Macchiavello and Morjaria (2015a) for details). Selling cherries to mills also yields higher revenues at the farm gate. The average price of cherries sold to mill was about 200 Rwandan Francs (RWF) per kilogram. In contrast, home processed parchment coffee fetched an average price of 760 RWF per kilogram. Since it takes approximately 5.5 to 6.0 kilograms of cherries to produce one kilogram of home-processed parchment irrespective of the processing method, the price of cherries under home processing is approximately 140 RWF per kilogram, substantially lower than the corresponding figure for cherries sold to mills.

The difference in prices underestimates the returns from selling cherries to mills for

[8]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. This step of the chain is not part of our analysis.

farmers since home-processed coffee entails additional processing costs for the farmers. As a result, farmers overwhelmingly report that selling cherries to mills is more profitable than home processing. For instance, when asked in our 2012 survey about which kind of buyer offers the highest price, only 2% of farmers answered traders buying home processed coffee. In a subsequent farmer survey in 2019 which also confirms the price difference, we asked farmers directly about the relative profitability of the two processing methods. Most of the farmers (98%) report that selling cherries to mills is more profitable than home processing.

Why do farmers engage in home processing at all given its much lower returns relative to selling cherries to mills? In both the 2012 and 2019 surveys, farmers reported that they would sell home processed coffee after the harvest period when they were in need of cash, effectively treating this production as a very expensive savings tool. This observation raises the question: why are mills unable to buy cherries at harvest and defer farmers' payments to the post-harvest period?

## 2.2 Mills and Farmers

*Survey:* To understand constraints to the operations of mills and farmers, we designed and implemented a survey in collaboration with the National Agricultural Exporting Board (NAEB) − the government institution in charge of the coffee sector. The survey was implemented towards the end of the 2012 harvest campaign (May through July) by four survey teams led by a qualified NAEB staff member.[9]

*Descriptive Statistics, Mills:* There are 214 processing mills in the country in 2012 (Online Appendix C, Figure C1). Summary statistics for mills in Rwanda are reported in Panel A of Table I. The survey covered all operating mills in the 2012 harvest season. The response rate was close to 100%.

The average mill employs around 35 seasonal employees and sources from close to 400 smallholder farmers. Coffee mills are thus large firms by developing countries' standards (see, e.g., Hsieh and Olken (2014)). There is dispersion in installed capacity, measured in tons of cherry processing per season. Small mills have capacity up to 250 tons; medium-sized mills, which constitute the majority, typically have a capacity of 500 tons; and a handful of large mills have a capacity in excess of 1000 tons.

Mills are characterized by a relatively simple technology that facilitates the calculation of unit costs of production. It takes approximately 5.5 to 6.0 kilograms of coffee cherries to produce 1 kilo of mill parchment coffee, the mill output. Under a Leontieff

[9]We complement our analysis with data from a farmer survey undertaken during the 2019 harvest.

technology approximation, the cost of producing 1 kilo of parchment coffee is the sum of ($i$) the price paid to farmers for cherries and ($ii$) other operating costs, including labor, capital, procurement, transport, marketing and overheads. The former accounts 60-70% of the total cost of processing.

*Descriptive Statistics, Farmers:* Summary statistics for farmers from the survey are reported in Panel B of Table I. The typical farmer is a smallholder who has completed primary education and owns a small coffee plantation of 500 to 1,000 coffee trees.

The sample of surveyed farmers was constructed as follows. When surveying a mill, we used a list of farmers from the coffee board's district office to randomly selected five farmers from the sector in which the mill is located.[10] The farmer survey is thus meant to be representative of all farmers located in sectors with mills, irrespective of whether or not the farmer sells to the mill.

We match our surveyed farmers in 2012 to a National Coffee Census conducted in 2009 to check whether our sample is representative of the population of farmers. We are able to locate the village of around 70% of the surveyed farmers in the census.[11] Online Appendix Table B1 compares our surveyed farmers in 2012 with those in the 2009 National Coffee Census. Within the relevant administrative sectors in which mills operate, farmers in the survey are similar to the wider population along a range of characteristics (household size, age, distance to the capital city, distance to the sector capital, distance to the nearest market trading centre as well as geo-physical conditions such as elevations, slope, Food and Agricultural Organization (FAO) coffee suitability conditions, presence of roads and rivers) but have more coffee trees and are closer to the mill.[12]

[10] Districts are the second-level administrative units in Rwanda. Sectors are the third-level administrative units, with an area of approximately 50 $km^2$. They are the lowest level at which the coffee board keeps regularly updated lists of active farmers.

[11] We only know the name of the farmer and the village where the farmer's plot is located. This would not per se be a major limitation given that the average village has an area just larger than 1 $km^2$. Unfortunately, however, village 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 a de-anonymized version of the national census of coffee farmers to assign farmers to a village and, thus, location. We are able to precisely locate approximately 70% of our surveyed farmers. We are able to locate an additional 10% of farmers through a fuzzy match procedure and find similar results when including those in our analysis. All results in our main Tables of the article are robust if we restrict the analysis to this restricted sample of farmers.

[12] The distance difference is likely due to the fact that the survey was conducted at the mill, and thus participation costs were higher for more distant farmers.

### 2.3 Relational Practices between Mills and Farmers

To operate efficiently, mills rely on relationships with farmers in the surrounding areas. Small holder farmers in developing countries typically lack access to well-functioning input and financial markets. Farmers resort to interlinked transactions (Bardhan (1989)) 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 I presents summary statistics for the main relational practices. We focus on practices for which the mill and the farmer exchange promises that are then fulfilled or reneged upon several weeks later; that is, those relationships 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 arrangement 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% (80%) of the farmers report to have received inputs (loans) from the mill (Table I, Panel B). The mill managers' survey yields similar figures (Table I, Panel A).

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 farmers and mills alike. As mentioned above, farmers report that a main motivation for home processing is to be able to sell coffee when they need cash rather than at harvest. Receiving part of the payments for cherries sold to mills during harvest as second payments paid after the end of harvest might thus help farmers overcome saving constraints. Mills might also benefit from purchasing cherries from farmers to reduce working capital requirements.[13]

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. Since farmers would provide trade credit in-kind (in the form of coffee cherries), input diversion on the part of the mill is unlikely to be the key concern (Burkart and Ellingsen (2004)). Farmers, however, might be concerned that the mill would renege on promised second payments. In the 2019 farmer survey we asked farmers whether they are concerned about mills defaulting on second payments. Of farmers that reported second payments, a third reported having experienced defaults on second payments in the past.

We asked managers whether the mill "has made second payments in the past" and farmers whether they "expect a second payment from the mill". The farmers' question captures the idea that, in relational contracting models, defaults occur off-the-equilibrium path. Concerns about default imply that promises of second payments might be constrained. On the extensive margin, the majority of managers and farmers report their use. Amounts typically are between 5% to 10% of total payments.

Finally, as part of the relational contract, the mill and the farmers can also transact services that are not related to harvest operations. For instance, 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 mulching trees ) or not (e.g., assistance 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. These qualitative dummies can be aggregated across the five surveyed farmers to measure how often farmers can rely on the mill for help/loans unrelated to harvest season. The provision of loans and inputs before harvest is also consistent with farmers' saving constraints ahead of the following harvest cycle.

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 in

[13] In the survey mills report that limited access to working capital finance is one of the main constraints to operation. This is consistent with evidence that coffee mills have large working capital requirements and are often credit constrained (Blouin and Macchiavello (2019)). For simplicity, in the theoretical section we model farmers saving constraints and abstract from mills credit constraints.

exchange for second payments; and finally, (*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 indices for the intensity of the relationship before, during, and after harvest giving equal weight to the managers' response and the average of the farmers' responses. Our main dependent variable is the overall "relational" contract index that aggregates the three period sub-scores.

There is significant dispersion in the adoption of relational practices. Figure I shows that the use of relational practices pre-harvest, at harvest and post-harvest are positively correlated across mills. The relational contract index thus captures a set of complementary relational practices. Online Appendix Figure C2 shows that the relational contract index correlates negatively with unit processing costs (panel A) and positively with capacity utilization (panel B). The relational contract index thus captures aspects of managerial practices that are appropriate to this industry.

## 3 Theory

This section lays out a theoretical framework that guides the empirical analysis. A mill interacts repeatedly with a population of farmers. In exchange for coffee cherries, the mill provides farmers with productivity-enhancing inputs and access to a saving tool through delayed payments. In the rural areas of developing countries, this type of intertemporal exchange is hard to enforce with formal contracts, and so the parties have to rely on relational contracts.

The model establishes two sets of results. First, we provide conditions under which competition between mills reduces parties' ability to sustain the relational contract. We derive predictions on mill- and farmer-level outcomes under these conditions. Second, we identify two distinct channels through which competition between mills affects relational contracts and we offer guidance on how to empirically disentangle them. The first channel, which we refer to as the *direct* effect, arises from the fact that after a farmer has already received productivity-enhancing inputs from a mill, she can choose to either deliver coffee cherries to that mill or sell them to an alternative mill. A larger number of competing mills makes this alternative more tempting, so that the original mill may be more reluctant to provide the farmer with productivity-enhancing inputs to begin with. The second channel, which we refer to as the *indirect* effect, arises from the fact that competition with other mills reduces a mill's profits. This will reduce the value of future rents, which are necessary to sustain the relational contract.

The model focuses on the most salient relational practices in our context: second payments, input extension and farmers' side-selling behaviour. We model second payments' role in alleviating farmers' saving constraints and the difficulty in enforcing them as they are critical in our context and not well emphasized in the literature. Besides empirical relevance, modeling input extension allows us to rationalize its complementarity with second payments despite the absence of a technological connection between the two. Finally, modeling side-selling offers a convenient way to tie temptations to deviate in the relational contract to the degree of competition.[14]

We model competition as a parameter that affects spot prices at which farmers can sell during the harvest season. This provides a parsimonious approach that still captures the direct and indirect effects of competition that are key to our analysis.[15]

## 3.1 Set-Up

*Players and Preferences*: A risk-neutral mill operates in an area populated by a unit mass of farmers, indexed $i \in [0, 1]$. Time is represented by an infinite sequence of identical seasons, indexed $t = 0, 1, 2..., \infty$. Within each season, there are three sub-periods, 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 $u(c_1, c_2) = \min\{c_1, c_2\}$. Consumption in each sub-period is equal to the sum of the transfer that the farmer receives from the mill and the revenue she earns from selling externally. These preferences capture farmers' demands for within-season consumption smoothing. The mill and the farmers have a common discount factor $\beta < 1$ across seasons. There is no discounting within season. In any season, the mill continues operation with probability $\theta$ (later endogenized) and ceases to operate with probability $(1 - \theta)$. Denote $\delta = \beta\theta$.

*Technology*: At harvest, farmer $i$ produces $Q_i^t$ units of coffee cherries. We describe $Q_i^t$ momentarily. Cherries must be processed at harvest. Once cherries are processed, they become 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 entails no additional cost. Home-processed coffee can be sold both at harvest and post-harvest at exogenous unit price $\rho$. Mill processing is performed by mills at constant marginal cost $c$. The mill

[14]We model input provision as affecting the volume of production, rather than its quality, and also abstract from farmers' heterogeneity in dimensions other than exposure to competition. Such extensions would match additional empirical findings.

[15]Both effects would also arise in a model in which entry is endogeneized and mills compete offering relational contracts. Such model introduces additional features that are not central to our analysis.

sells production at exogenous unit price $v$. As discussed in Section 2, mill processing is more efficient. We will make this precise in Assumption 2 below.

*Timing of Events*: Each season $t$ unfolds as follows (illustration of the timing is provided in the Online Appendix Figure C3):

> Pre-Harvest: (a) Mill draws a i.i.d. fixed cost $F^t \sim H(F^t)$; (b) Mill decides whether to pay the fixed cost and continue the game or exit. If the mill exits, the game ends and all parties get a payoff equal to 0; (c) Mill chooses whether to provide inputs to farmer $i$ at cost $k$, $\mathbf{I}^t_{k_i} \in \{0,1\}$.
>
> At Harvest: (a) Farmer $i$ harvests $Q^t_i = (1 + \mathbf{I}^t_{k_i}\pi)q$, $\pi > 0$ capturing increased yields from input extension; (b) Mill offers a payment $P^t_{1,i}$ in exchange for $Q^t_i$; (c) Each farmer decides whether to sell to the mill or not, $x^t_i \in \{0,1\}$.[16]
>
> Post Harvest: The mill decides whether to offer a second payment, $P^t_{i,2}$.

We make the following assumption:

**Assumption 1 (Contracts and Markets)**

a) The farmer does not have access to either input, credit or saving markets;

b) There is no formal contract enforcement: all promises must be self-enforcing.

The first part of the assumption introduces the motivation for interlinked transactions. The farmer lacks access to input, credit and saving markets. She needs to consume both at harvest and post-harvest. She can do that on her own through home processing, but that is inefficient. Alternatively, she can rely on the mill to get inputs to increase production and for savings through a post-harvest payment.[17]

The second part of the assumption, however, states that the mill's provisions of inputs and payments must be self-enforcing. Furthermore, the farmer's promise to sell to the mill after receiving inputs is also non-enforceable, capturing the well-documented side-selling problem in agricultural chains.

*The Relational Contract:* A relational contract between the mill and farmer $i$ is a plan that specifies $\mathcal{R}_i = \{\mathbf{I}^t_{k_i}, x^t_i, P^t_{i,1}, P^t_{i,2}\}^{\infty}_{t=0,1,\ldots}$ for all future seasons as a function of the past history of the game. We assume perfect public monitoring between the

[16] For simplicity, we assume that the mill either buys all cherries produced by the farmer or none. Results are qualitatively similar if farmers can sell a share of their harvest to the mill.

[17] The mill is assumed to have perfect access to the credit market.

mill and the farmer. A relational contract is self-enforcing if it constitutes a subgame-perfect equilibrium of the repeated game between the mill and the farmer.

We characterize the optimal relational contract that maximizes mill's profits. Specifically, before the beginning of season $t = 0$, the mill offers a relational contract to each farmer $i$ to maximize profits. Each farmer $i$ independently either accepts or rejects the offer, taking as given the actions of other farmers. If she rejects, both parties earn their outside option forever. If she accepts, parties enter the relational contract.[18] We focus on stationary relational contracts with grim-trigger punishment.[19]

Along an equilibrium path in which farmer $i$ sells to the mill in season $t$ ($x_i^t = 1$), the farmer's payoff is given by $min\{P_{i,1}^t, P_{i,2}^t\}$. The mill payoff conditional on operation and net of fixed costs $F^t$ is

$$\Pi^t = \int_0^1 (x_i^t((v-c)Q_i^t - P_{i,1}^t - P_{i,1}^t) - \mathbf{I}_{k_i}^t k)di \tag{1}$$

*Outside Options*: We now define outside options for the mill and for farmer $i$. In principle, there are two distinct outside options: before parties enter the relational contract; and following a deviation from either of the two parties after they have entered the relational contract. In both cases, we assume that parties stop trading with each other forever. This is also the case when the mill ceases operations.

The mill outside option is given by $u_m = 0$ since the mill does not process any of the farmer's coffee. To be precise, the mill sources coffee from other farmers. However, conditional on the mill operating, contracting and punishment are bilateral and therefore the mill's payoff from interacting with other farmers on- and off- the equilibrium path is independent of the relationship with farmer $i$.[20]

Farmer $i$'s outside option is defined as follows. If the farmer does not sell to the mill, she can sell cherries to other mills at harvest and home-processed coffee to traders at exogenous price $\rho$ at harvest and post-harvest.[21] Specifically, farmer $i$ can sell cherries at harvest to competing mills indexed $z \in \mathbf{C}_i \equiv \{1, ..., C_i\}$. $C_i$ ($\mathbf{C}_i$) is thus the number (set) of competing mills farmer $i$ can sell to. Competing mills buy from the farmer using spot contracts only. In a spot contract mill $z$ pays price $\rho_z$ at harvest and price

[18]When the farmer is indifferent, she is assumed to accept the offer. This rules out trivial coordination failures in which farmers reject simply because they think enough other farmers reject.

[19]A relational contract is stationary if $\mathbf{I}_{k_i}, q_i, P_{i,1}, P_{i,2}$ do not depend on $t$. When referring to stationary relational contracts we will drop the $t$ superscript.

[20]Farmers know the mill's exit probability (which depends on other farmers' actions) but are unable to coordinate punishment with other farmers scattered around the catchment area.

[21]We take the price for home processed coffee $\rho$ to be an exogenous parameter not affected by competition. The empirical analysis shows that competition between mills does not lead to higher prices for home processed coffee. This is likely due to free entry of traders with constant return to scale and an exogenous world price for home processed coffee at the export gate.

of zero post-harvest.

When selling to mill $z$, farmer $i$ faces iceberg transportation costs $(1 - \tau_{i,z})$. Denote with $z_i$ the mill that offers the best price net of transport costs and let $\rho_i = max_{z \in \mathbf{C}_i}\{\rho_z \tau_{i,z}\}$ denote such best price. The farmer's outside option is as follows. If $\rho > \rho_i$, the farmer home processes all her coffee and sells half of it at harvest and half post-harvest. This gives her payoff $Q_i^t \rho/2$. Otherwise, the farmer sells at harvest a share $(\rho/(\rho + \rho_i))$ of her production to mill $z_i$ as cherries at price $\rho_i$. She home processes the remaining share of her production and sells it post-harvest for price $\rho$. This gives her payoff $Q_i^t \times (\rho\rho_i/(\rho + \rho_i))$.

The farmer's payoff in the outside option is thus equal to $Q_i^t \times \rho \times (max\{\rho_i, \rho\}/(\rho + max\{\rho_i, \rho\})) = Q_i^t \times u(\rho, C_i)$ which is (weakly) increasing in the price for home processed coffee $\rho$ and in the number of competing mills $C_i$ the farmer can sell to. The reduced-form outside option $u(\rho, C_i)$ captures the idea that, in the absence of a relational contract with the mill, the farmer sells at least part of her production as home processed coffee post-harvest in order to save. The remaining part of her production will be sold at harvest either as cherries or as home-processed coffee. The value of the outside option (weakly) increases with exposure to other mills, $C_i$, which is assumed to vary across farmers depending on their location.

We make the following assumption:

**Assumption 2 (Technology)**: $(v - c) > u(\rho, C_i)$ and $k < q\pi(v - c)$.

The first part of the assumption captures the fact that mill processing is efficient (see the discussion in Section 2). The second part of the assumption states that pre-harvest inputs given by the mill are also efficient: they increase joint surplus by $q\pi(v - c)$ and only cost $k$.

*Incentive Compatibility Constraints:* We derive conditions under which the following actions occur in each period of a stationary relational contract: $(i)$ at pre-harvest either $\mathbf{I}_{k_i} = 1$ or $\mathbf{I}_{k_i} = 0$; $(ii)$ the farmer sells to the mill, $x_i = 1$; and $(iii)$ the mill makes payments $P_{i,1}$ and $P_{i,2}$ at harvest and post-harvest.

The two key incentive compatibility constraints are the ones ensuring that the farmer doesn't side-sell and that the mill pays second payment $P_{i,2}$.[22]

At harvest, the farmer must prefer to sell to the mill rather than side-sell and then losing access to the mill in the future. The farmer's per-period payoff in the relational contract is given by $u(c_{i,1}, c_{i,2}) = \min\{P_{i,1}, P_{i,2}\}$. If the farmer side-sells she

[22]The incentive compatibility constraints associated with input provision and payment of $P_{i,1}$ are slack. Details of all incentive constraints and proofs are provided in the Online Appendix A.

gets $(1+\mathbf{I}_{k_i}\pi)q \times u(\rho, C_i)$ this season and her outside option $q \times u(\rho, C_i)$ forever after. This gives the no side-selling incentive constraint:

$$min\{P_{i,1}, P_{i,2}\} + \frac{\delta}{1-\delta} min\{P_{i,1}, P_{i,2}\} \geq (1+\mathbf{I}_{k_i}\pi)qu(\rho, C_i) + \frac{\delta}{1-\delta} qu(\rho, C_i) \quad (2)$$

Post-harvest, the mill must prefer to pay the second payment $P_{i,2}$ and continue the relationship rather than defaulting and obtaining her outside option equal to zero from then onward. The incentive constraint is given by:

$$\frac{\delta}{1-\delta}((v-c)(1+\mathbf{I}_{k_i}\pi)q - P_{i,1} - P_{i,2} - \mathbf{I}_{k_i}k) \geq P_{i,2}. \quad (3)$$

*Which farmers can sustain the Relational Contract?* The relational contract maximizes the mill profits, and thus it must be that $P_{i,1} = P_{i,2}$ and that (2) is binding. This implies

$$P_{i,1} = P_{i,2} = ((1-\delta)(1+\mathbf{I}_{k_i}\pi) + \delta) \times qu(\rho, C_i) \quad (4)$$

Substituting (4) into (3) we obtain the necessary condition under which a self-enforcing relational contract exists. The condition is given by:

$$\frac{\delta}{1-\delta}((v-c)(1+\mathbf{I}_{k_i}\pi)q - \mathbf{I}_{k_i}k) \geq u(\rho, C_i)q(1+(1-\delta)\mathbf{I}_{k_i}\pi) \quad (5)$$

This condition states that the net present value of the per period rents generated by selling cherries to the mill (given by $(v-c)(1+\mathbf{I}_{k_i}\pi)q - \mathbf{I}_{k_i}k$)) ought to be larger than the aggregate temptation to deviate (which is equal to the second payment $P_{i,2}$).

## 3.2 The Direct and Indirect Effects of Competition

Condition (5) gives:

**Proposition 1 (Direct Effect of Competition)**

1) *For each farmer $i$ there exist unique thresholds $\delta_i^{\mathbf{I}_{k_i}}$ such that if $\delta \geq \delta_i^{\mathbf{I}_{k_i}}$ a self-enforcing relational contract between farmer $i$ and the mill with $\mathbf{I}_{k_i}$ exists*;

2) *The two thresholds $\delta_i^{\mathbf{I}_{k_i}}$ are increasing in farmer $i$ exposure to competition $C_i$*;

3) *If $(v-c)q < k/\delta$ relational contracts with post-harvest payments but no input provision are never sustainable.*

The first statement in Proposition 1 follows standard logic: a self-enforcing relational contract exists if the discount factor is sufficiently large.

The second statement in Proposition 1 gives us the direct effect of competition. The right hand side of condition (5) is increasing in farmer $i$ exposure to competition $C_i$. All else equal, farmers with higher access to competing mills will find it harder to sustain the relational contract with the mill than farmers with lower access to competing mills.

Finally, the third statement in Proposition 1 states that, under certain conditions, relational contracts with post-harvest payments but no input provision cannot be sustained.[23] When this happens, relational practices are complementary in the sense that they move together with a change in the underlying competition parameter (see Brynjolfsson and Milgrom (2013)).

Competition can also have an *indirect* effect on the relational contract between the mill and farmer $i$. Recall that $\delta = \beta\theta$, with $\beta$ the common discount factor between parties and $\theta$ the probability that the mill continues operations. We now endogenize $\theta$.

In the stationary equilibrium the mill's variable profits are constant over time and increasing in the (mass of the) set of farmers with whom the mill sustains a relational contract. Denote such set $i \in \mathcal{R} \subset [0, 1]$. We have

$$\Pi = \int_{i\in\mathcal{R}} (q(v - c)(1 + \pi\mathbf{I}_{k_i}) - \mathbf{I}_{k_i}k)di \tag{6}$$

At the beginning of every season $t$ the mill draws fixed costs $F^t$ from the cumulative distribution $H(F^t)$. The draws are i.i.d. over time. Upon observing the fixed costs $F^t$ the mill decides whether to pay the fixed costs and continue operations or not, in which case it exits the market and all relationships with farmers come to an end.[24]

The mill exits when fixed costs $F^t$ are above a threshold $\overline{F}(\Pi)$ increasing in $\Pi$. The probability that the mill continues operation, $\theta$ is given by $\theta = H(\overline{F}(\Pi))$ and is thus increasing in $\Pi$.

Mill-level competition is defined as the union of the sets $\mathbf{C}_i$, that is, $\mathbf{C} = \cup_{i\in[0,1]}\mathbf{C}_i$.

**Proposition 2 (Indirect Effect of Competition)**: *Consider an increase in mill-level competition* $\mathbf{C}$ *induced by an expansion in the sets* $\mathbf{C}_i$ *for a positive mass of farmers* $i \in \mathcal{R}^C \subset \mathcal{R}$. *Suppose the increase in competition destroys the relational contract through its direct effect for a positive mass of farmers* $i \in \mathcal{R}^C$. *Then the*

[23]This is because input provision has an ambiguous effect on the sustainability of a relational contract. On the one hand, it increases joint profits and thus makes it easier to enforce a relational contract. On the other hand, it also increases the farmer current outside option and the second payment $P_{i,2}$, making it harder to sustain a relational contract.

[24]Exit means that the current owner sells the mills to a new owner and relational contracts in place end. Changes in ownership are not uncommon in the industry (see Macchiavello and Morjaria (2020)).

*relational contract might no longer be an equilibrium for some other farmers $j \in \mathcal{R}$, $j \notin \mathcal{R}^C$.*

Proposition 2 follows from the fact that an increase in competition that destroys the relational contract for a positive mass of farmers leads to a decrease in the mill's processed volume and thus in variable profits $\Pi$. This lowers the probability that the mill continues operation, $\theta$, and, therefore, $\delta$. By the first statement in Proposition 1, this can further destroy the relational contract for farmers $j \in \mathcal{R}$ that were not directly affected by the original increase in competition (that is, $j \notin \mathcal{R}^C$). This is the indirect effect of competition of relational contracts.

The indirect effect of competition might kick in only when there are sufficiently many competing mills. The impact of an additional competing mill on the sustainability of relational contracts between the mill and surrounding farmers might become stronger as competition intensifies.

When competition destroys relational contracts, the model delivers a cluster of predictions on additional mill-level and farmer-level outcomes.

First, higher competition $C_i$ can make the farmer worse off when it destroys a relational contract with $\mathbf{I}_{k_i} = 1$. Since the side-selling constraint (2) is binding, the farmer's utility in a relational contract with $\mathbf{I}_{k_i} = 1$ is strictly higher than the utility under no relational contract or in a relational contract with $\mathbf{I}_{k_i} = 0$. Note that farmers can be worse off even when competition increases prices for cherries at harvest. Due to the lack of saving tools, a farmer cares about when she is paid, not only how much.

Second, when the relational contract cannot be enforced there is no spot price at which farmers sell all their production as cherries 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, the quantity of cherries sold for processing at harvest declines at the farmer, at the mill and at the aggregate level. Given fixed costs and constant variable processing costs, the lower quantity processed by the mill also implies higher average cost.

An extension of the model in which farmers also exert costly, non-contractible, effort yields that prices paid at harvest could also decrease due to competition. In such an extension, the price paid by the mill must compensate farmers for the effort and, if competition makes it impossible to sustain effort, observed prices paid by the mill might also decrease. The impact of competition on prices is thus ambiguous. In a similar vein, if the farmer's non-contractible effort increases the quality of the coffee, competition can lower the quality of coffee produced by the mills.

### 3.3 Summary of Predictions

We summarize the predictions of the model as follows:

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

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

  (1) The use of all relational contract practices (inputs pre-harvest, second payments and help/loans to farmers) decreases;

  (2) Mills process lower volumes of cherries, have higher average costs, and produce lower quality;

  (3) Farmers sell fewer cherries at harvest to any mill, have lower revenues and are worse off;

  (4) Prices paid to farmers at harvest may increase or decrease;

**C.** Competition reduces relational contracts with farmers that can directly sell to the competing mill as well as, indirectly, with farmers that cannot sell to those mills.

These predictions are empirically tested in the rest of the article. Section 4 asks whether competition breaks relational contracts (prediction **A**). Section 5 explores the consequences of relational contracts' breakdown for mills and farmers (prediction **B**) while Section 6 tests for mechanisms (prediction **C**).

## 4 Does Competition Break Relational Contracts?

### 4.1 Measuring Competition

We take a conservative approach and define the catchment area to have a 5 km radius.[25] Two mills compete with each other if their catchment areas overlap. Given this definition, the baseline measure of competition is the number of mills within a 10 km radius from the mill (see Online Appendix Figure C4 for an illustration).

There is significant dispersion in the intensity of competition faced by mills (see Online Appendix Figure C5). 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.

[25] On average, mills' managers report catchment areas with a radius of $\approx$ 4.5km.

The survey asked the mill's manager the number of other mills that source coffee cherries inside the mill's catchment area. The average manager reported competition from about 6 mills 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, however, are heterogeneous with respect to both installed capacity and density of coffee trees in their catchment area. 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 might endogenously respond to both competition and to mill's practices. Mill 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 simplicity, we present OLS and IV results using the baseline measure and discuss robustness checks that use mill-specific measures of competition in Online Appendix D.

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

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

$$RC_m = \alpha + \beta C_m + \eta X_m + \gamma Z_m + \varepsilon_m, \tag{7}$$

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, cooperative status and mill coordinates). The vector $Z_m$ includes geographic controls for potential drivers of the mill's performance within the mill's catchment area: elevation, slope, presence of spring, density of coffee trees, length of roads and rivers and coffee suitability from FAO's Global Agro-Ecological Zones (FAO-GAEZ).

Column 1 of Table II shows that competition negatively correlates with relational contracts: an additional competing mill is correlated with a 0.116 standard deviation lower relational contract 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 (the first stage), and (*ii*) does not influence the use of relational contracts with farmers other than through its effect on competition (the 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"). Conditional on suitability within the mill's catchment area, competition is instrumented with suitability in the adjacent area 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 radius from the mill, conditional on suitability (and other controls) within the 5 km radius catchment area.

We build on 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 GIS specialists developed an engineering model for the optimal placement of mills in Rwanda (see, Schilling and McConnell (2004)). Given the particularly rugged nature of Rwanda, the model intended to identify suitable sites for mill construction at a high spatial resolution taking into account a vector of characteristics to be then aggregated into a "suitability score". The model, however, was never fully implemented because the required GIS data were not readily 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 25 $m^2$ resolution we run the engineering model for the whole of Rwanda at a resolution of 1 $km^2$.

The engineering model specified four criteria for a mills' placement: (1) mills should be outside National Parks, Natures Reserves and other protected and conservation areas; (2) in sectors with at least 30,000 coffee trees; (3) Within 3 km from a spring source, at an elevation between $-10$ meters and $-30$ meters from the spring; and (4) within 1 km of a road. For each 1 $km^2$ square in Rwanda (henceforth, "grid") we define dummies for whether it satisfies each of these four criteria or not. Online Appendix Figure C6 illustrates spatial variation in the engineering model's criteria.

We build on the engineering model and construct our instrument as follows. There are thousands of potential grids where mills could have entered and 214 in which a mill had entered by 2012. All mills that have entered satisfy criteria 1 and 2. Grids not satisfying these two criteria are thus assigned a suitability score equal to zero. Within the sample of grids satisfying criteria 1 and 2 we run a probit model to predict mill entry. The probit model includes dummies for the remaining criteria 3 and 4, their interaction, and additional controls (polynomials in distances to springs and roads, average elevation and slope in the grid, longitude and latitude of the grid box centroid, density of coffee trees in the grid box, size of the sector and interactions of these variables).

The probit model lends support to the engineering model. Online Appendix Table B2 shows that the interaction between dummies for criteria 3 and 4 predicts mill's placement (p-value $< 0.01$). We use estimates from column 4 to predict a suitability score for each 1 $km^2$ grid (Online Appendix Figure C7 illustrates the results). Finally, we aggregate the predicted suitability scores at the mill level. The average suitability score in the grids within a 5 km radius from the mill gives us a control for suitability in the mill's catchment area. Our instrument is the average suitability score within the area of 5 to 10 kms radius from the mill, akin to a cross-section surface of a donut.

The engineering model criteria and the controls raise a number of concerns about the identification strategy. First, coffee trees and roads inside the catchment area could be endogenous. This could generate a "bad control" problem. Second, conditional on these controls, trees and roads outside the mill catchment area could influence mill and farmers inside the catchment areas through channels other than competition (a violation of the exclusion restriction). We first present our main result and then present robustness checks that address these threats to our identification strategy and also explore robustness of our results along other dimensions.

### 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 radius from the mill. Specifically, the first stage is given by

$$C_m = \alpha + \widehat{\beta} S_m^{5/10} + \beta S_m^{0/5} + \widehat{\gamma}_0 X_m + \widehat{\gamma} Z_m + \mu_m \tag{8}$$

where $S_m^{5/10}$ is the average predicted engineering model suitability score in the donut area between 5 and 10 km from mill $m$, $S_m^{0/5}$ is the average predicted engineering model suitability 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 (7). The exclusion restriction is satisfied if, conditional on suitability 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.

Panel A of Figure II shows a strong first stage: the predicted score $S_m^{5/10}$ strongly correlates with competition $C_m$. Column 2 in Table II reports the results. An increase of one standard deviation in the instrument $S_m^{5/10}$ is associated with mill $m$ facing competition from 1.610 additional mills (p-value $< 0.01$).

Panel B of Figure II shows a strong reduced form relationship between the instrument, $S_m^{5/10}$, and the relational contract index, $RC_m$. Column 3 in Table II reports the estimates. A one standard deviation increase in the instrument $S_m^{5/10}$ is associated with a reduction of 0.455 standard deviations in the relational index (p-value $< 0.01$).

Column 4 in Table II reports the 2SLS estimates. An additional mill within a 10 km radius from the mill causes a reduction of 0.283 standard deviations in the relational contract 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 more than twice as large as the OLS (-0.116 vs -0.283). 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.

The specification assumes a linear effect of the number of competing mills on the relational contract index. In the model 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 expect the negative effect of competition to become stronger as competition intensifies, at least up to a certain point. Online Appendix Figure C8 explores the functional form of the relationship between competition and relational contracts reporting results from non-parametric IV estimation. The estimates indeed 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. This pattern is consistent with the predictions of the model in Section 3.

### 4.5 Threats to the Identification Strategy

We now discuss threats to the identification strategy. First, we consider the role of both the presence of roads and tree density as ingredients of the instrument as well as their potential role as bad controls. We then consider several mechanisms that could lead to a violation of our exclusion restrictions. Online Appendix D further explores robustness along other dimensions, including the definitions of competition and catchment areas and other potential threats to the identification strategy.

#### 4.5.1 Exclusion Restriction and Bad Controls: Roads and Coffee Trees

The logic of the identification strategy is that, conditional on road access and coffee tree density inside the mill's catchment area, roads and coffee tree density in the donut area do not directly affect farmers and mills. The logic is potentially undermined by two distinct sets of concerns. First, coffee tree density levels and road access outside the catchment area could affect mills and farmers directly. For example, a road in the donut area could still be used by the mill or by farmers; prices for home processed coffee might depend on harvest levels in the donut area through general equilibrium effects, and so on. If that is the case, the exclusion restriction is violated. Second, if road construction and coffee tree density in the catchment area respond to mills' operations, conditioning on these variables in the catchment area could induce a bad control problem.

Table B3 in the Online Appendix investigates the robustness of our baseline results to these concerns. Column (1) reports, for ease of comparison, our baseline specification. Columns (2) and (3) considers road presence. Column (2) removes roads from the IV, that is, from the engineering model used to predict the suitability score inside the donut area. This addresses the concern about the violation of the exclusion restriction, but not concerns about roads being a bad control. Column (3) thus goes one step forward and removes roads from the construction of the IV and the suitability score inside the catchment area, as well as a control. In both cases, results are robust: both the first stage (reported in Panel B) and the second stage (Panel A) remain highly statistically significant and has a similar magnitude as the baseline in column (1).

Columns (4) through (8) consider coffee trees. First, the second criteria of the engineering model restricts the sample of suitable grids to those in sectors (administrative unit of Rwanda) with the presence of a least 30,000 coffee trees. Column (4) replaces this criterion in the entry model presented in the Online Appendix Table B2 with a restriction requiring that the grid has a suitability for coffee cultivation (from FAO-GAEZ) equivalent to at least 460 tonnes per hectare. Results are virtually unchanged.

Column (5) removes tree density from the construction of the instrument. Analogously to column (2) for roads, this is meant to address concerns about violations of the exclusion restriction. The specification thus leaves only suitability for coffee from FAO-GAEZ as an indicator of coffee activity in the entry model predicting suitability of mill placement. Results are again essentially unchanged. We interpret the similarity between the baseline specification in column (1) and the results in columns (2) and (5) in the spirit of an over-identification test: as our instrument relies on multiple sources of variation, we can construct alternative instruments exploiting only subsets of these sources and obtain similar results.

Column (6) further removes tree density as a control and from the construction of the suitability score inside the catchment area. The magnitude of the second stage point estimate drops by about a third (from -0.283 to -0.182). Although the first and second stages are still significant, this suggests that tree density in the catchment area could be a bad control.

While tree density in the mill's catchment area might respond to the mill's entry and operations and thus can be a potential "bad control", it seems a priori important to control for it. For example, the same level of competition could have different impacts on the mill depending on how much coffee is grown in the region around the mill. To this end, FAO-GAEZ suitability for coffee cultivation is not a sufficiently precise control for two reasons: (*i*) it is defined at a much higher level of aggregation (at the 9 $km^2$ resolution) than our analysis and is thus weakly related to variation in local conditions; (*ii*) there are places that are suitable for coffee cultivation but are occupied by other economic activities (e.g., urban developments, conservation zones and mines).

Columns (7) and (8), therefore, repeats the exercise computing coffee tree density using the National Coffee Census conducted in 1999; at that time, only two mills had been built in Rwanda. Tree density in 1999 is thus not the result of subsequent mill entry by the time of the survey in 2012. Although changes in administrative boundaries introduce measurement error, this strategy nearly halves (from −0.182 to −0.224) the gap in the point estimate relative to the baseline (−0.283).

A potential concern is that both roads and tree density feature exclusion restriction violations. In such case, the results could remain (erroneously) robust when retaining at least one of the two in the calculation of the instrument. Column (9) reports results from a specification in which both tree and road density have been removed form the construction of the instrument and show that the results are robust.[26]

### 4.5.2 Farmers Outside Options

Although the results are robust to the exclusion of roads and coffee tree density from the instrument, it is still possible that suitability for mill entry in the donut area around the mill catchment area affects farmers' and mills' operation through channels other than competition for coffee. For instance, the instrument could correlate with better access to, or wages in, work outside the coffee sector; access to alternative saving or other financial services beyond the mill; and convenience of, or price available for, post-harvest sales of home-processed coffee or other crops. In all these cases, the demand for interlinked transactions with the mill is lower for reasons unrelated to competition between mills.

We directly check for these potential exclusion restriction violations in the data. We first consider whether our instrument correlates with farmer characteristics that should not be affected by coffee production or sales. We then check whether our instrument correlates with economic opportunities for farmers. Finally, we explore whether the results are robust controlling for farmers' proxies for market access. In all these cases, we use farmer-level specifications that include farmer's age, gender, place of birth, education level, cognitive skills, distance from the mill and the farmer's coffee tree holdings as additional controls.

Online Appendix Table B4 explores farmer characteristics from our 2012 survey (age, gender, schooling, cognitive test) as well as from the 2009 National Coffee Census (household size and age). Panel A finds no correlation between our instrument and farmers' demographic characteristics. Panel B finds some correlation between competition and farmers' demographics.

[26]As a further robustness test, column (10) shows that estimating the engineering entry model at the grid level with an OLS rather than with a probit model yields nearly an identical first stage and slightly larger second stage point estimates.

Online Appendix Table B5 explores the correlation between our instrument and measures of farmers' outside economic opportunities. Unfortunately, our 2012 farmer survey did not include much information on farmers' economic activities outside of coffee production. We conducted a representative survey of farmers in the 2019 season to gather direct evidence on the extent to which our instrument is correlated with better outside options and/or access to financial services.

The results support our exclusion restriction. We find that the instrument does not correlate with the percentage of coffee income in the farmer's total income (column 1); the likelihood the farmer has other sources of income (column 2); the likelihood of the farmer being employed by others both on the extensive (column 3) and the intensive (column 4) margin; conditional on employment, the wage rate (column 5) and the total wage income (column 6); the payment due to employing additional labor on the farm (column 7); the likelihood the farmer sells milk (column 8); conditional on selling milk, the price and amount of milk sold (columns 9 and 10); access to formal saving accounts from Banks and/or local saving cooperatives (columns 11 and 12). This survey evidence suggests that our instrument does not correlate with economic opportunities that might lower the demand for relational contracts with the mill and thus supports the validity of our exclusion restriction.

Online Appendix Table B6 considers an alternative strategy that controls for farmers' market access. A potential concern with the results in Online Appendix Table B5 is that the survey evidence was gathered in the 2019 season, seven years after our baseline evidence. For farmers for whom we have exact location information in 2012, we construct measures of market access along the lines suggested by Donaldson and Hornbeck (2016) using urban population (from the 2012 National Population and Housing Census) and the most detailed road infrastructure data we have from the 2008/09 aerial ortho-photos. For ease of comparison, columns (1) and (2) reports our baseline results on the full sample of farmers and on the sample of farmers for which we have exact village location information respectively. Columns (3) through (6) include measures of market access, defining markets relative to any of the 62 officially designated urban centres (weighted by population); sector capital (weighted by population); to the capital city (Kigali); and lastly all official market trading centres. Results are robust across all these specifications.

# 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 mill-level and farmer-level outcomes. This section tests these additional predictions.

## 5.1 Complementarties in Practices (Prediction B1)

The model implies that relational practices might be complementary. Competition alters only farmers' ability to sell cherries at harvest to a competing mill. When competition increases, however, all practices might become unsustainable. Table III reports OLS (Panel B) and IV (Panel A) specifications considering relational practices one at a time. For each practice, the Table reports specifications using farmers' responses, managers' responses, and the aggregate of the two.

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 use of this practice. Aggregating farmers' and managers' answers, we find that competition from an additional mill reduces the use of this practice by 0.220 standard deviations (column 3).

Columns 4 to 6 ask whether competition reduces sourcing of cherries on credit at harvest for which the mill pays second payments to farmers. Regardless of whether we ask farmers or managers, competition causes a reduction in use of this practice. When answers from farmers and managers are aggregated, competition from an additional mill reduces the use of this practice by 0.203 standard deviations (column 6).

Finally, columns 7 to 9 ask whether competition reduces assistance and help to farmers post-harvest. Competition from an additional mill reduces the use of this practice by 0.180 standard deviations (column 9). Column 10 aggregates the three relational contract practices by respondents and creates an index. The relational contract indices are also separately reported by respondent type (columns 11 and 12).

The model focuses on relational practices for which lack of contract enforcement matters, i.e., those in which the mill and the farmer exchange non-enforceable promises across several weeks. In contrast, we expect lack of contract enforcement to be less of a concern for exchange of promises over very short periods. We consider short-term credit and advances during harvest, two practices driven by liquidity considerations and that are not part of the relational contract between the mill and the farmer. Results in column (13) confirm that competition does not impact this type of short-term credit between the mill and the farmers.

In sum, the evidence supports the idea that relational practices are complementary: competition reduces the use of all relational contract practices simultaneously.

### 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 IV investigates these predictions. Column 1 shows that unit costs of processing 1 kilo of the output increases by 4.6% with an additional competing mill. Columns 2 and 3 show no effect on prices for cherries paid to farmers during harvest nor on the price for home processed parchment in the area, as reported by the mill manager. Column 4 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 4 implies that competition has no effect on the cost of cherries. The cost of cherries accounts for about 60% of the overall unit costs of output production at the typical mill. The coefficient in column 1 must thus be explained by increases in other operating costs. Accordingly, column 5 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 6 shows that competition reduces the total volume of coffee cherries processed by the mill. An additional competing mill is associated with approximately 5 fewer tons of processed cherries. This translates into 7.5% lower capacity utilization, a sizeble effect given that the average capacity utilization in the industry is around 50%.

The breakdown in relational contracts with farmers makes deliveries harder to plan for. Column 7 shows that competition does not affect the number of weeks the mill is in operation during the harvest. Competition, however, increases the likelihood that the manager reports to have had both days with too many and too few workers at the mill (columns 8 and 9).[27] The difficulty in planning results in higher labor costs. Column 10 shows that the labor component of unit costs increases with competition:

[27] Labor costs increase as mills do not perfectly adjust labor to irregular deliveries. While 65% of mills revise employment plans weekly depending on cherry procurement and market conditions, 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 to process. For example, 73% (12%) of mill managers report that they would turn down only some (none) of the workers if there were few cherries to process. Mills are located in densely populated rural areas with few employment opportunities and so competition has no impact on wage rates nor on the manager reporting difficulties in hiring workers.

an additional mill increases unit labor costs by nearly 11%.[28]

*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 might 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 be processed.

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. The cupping process scores each sample along several dimensions of quality related to both physical characteristics of the processed coffee (parchment) 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 V 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 quality 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: parchment 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 0.172 standard deviations.

We also construct an index that captures quality dimensions that are mostly influenced by sorting and drying practices at the mill. The index aggregates moisture content, floating beans and broken beans as dimensions of quality. Column 3 shows no impact of competition on the index of mill-related practices. 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.

In sum, the evidence is consistent with competition increasing mills' operating costs and reducing the quality of the coffee produced through its negative impact on relational practices with farmers.[29]

[28] Capital, transport, and procurement are the main other sources of operating costs. We do not find significant effects of competition on these other costs.

[29] In the Online Appendix Table B7 shows that the main mill-level results are robust to the main

### 5.3 Farmer Outcomes (Prediction B3)

The model predicts that a breakdown in the relational contract with the mill is associated with the following changes in farmer-level outcomes: (*i*) an ambiguous effect on prices paid to farmers, (*ii*) a drop in the share of cherries sold to mills (since farmers cannot rely on the mill's second payments to smooth cash flows), (*iii*) a reduction in access to inputs, and, finally, (*iv*) lower revenue and welfare.

Table VI tests these predictions with farmer-level specifications. Column 1 confirms the finding of column 2 in Table IV: competition has at best a small 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 around 1%. Note that this is the price farmers report for sales of cherries during harvest. Since competition reduces second payments after the end of harvest, this estimate provides an upper bound to the effect of competition on the net-present-value of payments to farmers. Furthermore, column 2 shows that competition between mills does not change prices received for home processed parchment coffee. This result, in line with column 3 of Table IV, confirms that the impact of competition on prices received by farmers is negligible and supports our approach to model the price of home processed parchment as an exogenous parameter.[30]

Column 3 shows that competition reduces the share of a farmer's production sold as cherries to any mill during harvest. That is, competition between mills actually increases the share of coffee that is home-processed. Column 4 finds that competition increases the likelihood that farmers report saving as the main motivation for processing coffee at home rather than selling cherries at harvest to the mill. 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 their coffee income after harvest. Competition destroys the relational contract between the farmer and the mill, in particular the mill's ability to credibly promise payments after the harvest. Farmers are then forced to process part of their coffee at home in order to save income until after harvest.[31]

Column 5 shows that competition increases the likelihood that farmers have to

robustness checks performed in Table B3 and to the alternative measure of competition in Table B8.

[30] Since competition reduces quality, we might potentially underestimate the quality-adjusted price received by farmers. We think this is unlikely because mills did not pay farmers based on quality. Limited quality price premia at the farm-gate are not specific to our context (see e.g., Minten et al. (2018) for Ethiopia, Macchiavello and Miquel-Florensa (2019) for Colombia, Morjaria and Sprott (2018) for Uganda).

[31] The results show that the aggregate amount of cherries sold by farmers to mills decreases as a result of competition. We discuss this further at the end of Section 6 and in the Online Appendix B

self-finance inputs without an increase in yield (measured as kgs of cherries per coffee tree, column 6) nor an increase in overall input usage (measured as RWF spent per kg of cherries, column 7). Column 8 also shows that competition does not lead farmers to invest in their plantation and increase the number of coffee trees.[32]

The lack of an effect of competition on prices, yields and input usage suggests that competition does not increase farmers' returns from coffee cultivation. It is notoriously difficult to measure profits for farming enterprises.[33] We nevertheless compute overall revenues from coffee cultivation adding reported sales of home parchment and cherries sold to mills. Column 9 finds that competition reduces revenues by 8% (p-value <0.10). This is due to no change in overall production, a higher share sold as home processed parchment coffee, and the lower prices fetched by home processed coffee. The estimated impact on revenues likely understates the negative impact on farmers profits and welfare since (*i*) holding prices constant, farmers have to save through a very costly mechanism (home processing); (*ii*) farmers incur higher costs in order to home process coffee.

Given difficulties in measuring revenues and profits, we also consider the effect of competition on an overall index of job satisfaction as our preferred proxy for farmers' welfare. Column 10 in Table VI shows that competition has a strong negative impact on farmers' overall reported satisfaction. Columns (11), (12) and (13) open up the job satisfaction index and finds that competition lowers the likelihood that the farmer reports that the pay from the coffee business is good, further supporting our results on income. Therefore, the evidence supports the model's predictions on farmer-level outcomes and suggests that farmers might not benefit from competition.[34]

# 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 a given farmer. First, there is a direct

[32]In contrast to the model's prediction, we do not find evidence that competition lowers the volume of production.

[33]These difficulties are particularly pronounced in our context as (*i*) farmers' have low literacy levels; (*ii*) coffee cultivation coexists alongside several other farming and non-farming activities; (*iii*) we implemented our survey before the end of the harvest season; (*iv)* the length of the farmer survey we could implement at the mill was severely constrained.

[34]In the Online Appendix columns (7) to (12) of Table B6 shows that the main farmer level results are robust to controlling for market access. Similarly, Table B7 shows that the main farmer-level results are robust to the main robustness checks to the exclusion restriction performed in Table B3 and to the definition of catchment area performed in Table B8.

temptation mechanism: when competition increases the farmer's outside option it becomes harder to sustain the relational contract between the mill and the farmer. Second, there is an indirect profit mechanism: competition reduces the mill's profits and likelihood of operating in the future and thus makes it harder to sustain a relational contract even with farmers not directly affected by competition.

Table VII untangles the two mechanisms. The intuition is as follows. Holding constant mill-level competition and farmer's distance to the mill, the nearest is the farmer to competing mills, the higher is the farmer's outside option. We thus expect that proximity to competing mills is correlated with a lower relational contracting index between the mill and the farmer. To explore this hypothesis, we need to compute distances between each farmer and all mills. We are able to do so for 70% of the surveyed farmers (see footnote 11).

Column (1) reports for convenience our baseline specification at the farmer level. Column (2) repeats the exercise on the sample of farmers that we can match to an exact location and for which we can compute distance to mills. Column (3) then adds a measure of farmer specific mill access to our baseline farmer-level specification. Analogously to Donaldson and Hornbeck (2016), mill access is constructed as the inverse of the sum of the distance to the nearest and second nearest competing mills to the farmer.[35] We find that, conditional on mill-level competition, a farmer's proximity to competing mills is correlated with a lower relational contract index.

A potential concern is that the farmer's mill access might be endogenous. To assuage such concern, Column (4) repeats the exercise only including the inverse of the distance to the second nearest competitor and finds similar results. Ideally, however, we would like to instrument for the farmer's access to mills following an IV strategy similar to that used for mill-level competition. To do so, we would like to construct an instrument for the suitability of mills in a donut area around the farmer. Such an approach yields instruments for mill-level competition and for farmer-level access to mills that are strongly correlated with each other.

We therefore pursue an alternative strategy in the remaining columns of Table VII. The approach relies on dividing the area surrounding the mill into four "quadrants" (that is, quarters of a circle): north-west, north-east, south-east, and south-west. Each farmer is then assigned to her quadrant. For each farmer we split competition into the number of mills in the farmer's quadrant (hence, farmer competition) and in the three other quadrants (mill competition).

[35] We take exponent $\epsilon = 2$ but results are robust to alternative $\epsilon$. This approach parallels the robustness checks on farmer market-access in the Online Appendix Table B6.

Conditional on suitability in the farmer's quadrant of the mill's catchment area, we instrument competition in the farmer's own quadrant with the average suitability score in the relevant portion of the 5-10 km donut. We construct an instrument for competition from mills in other quadrants in the same way, controlling for average suitability in the remaining quadrants of the mill catchment area.

The approach relies on the idea that competition from mills in other quadrants only affect farmers through the indirect profit mechanism. In contrast, competition from mills in the farmer's quadrant affects the farmer both through the direct temptation mechanism as well as through the indirect profit mechanism. Although the strategy only proxies for the two distinct channels through which competition operates, the two instruments are computed on different regions of the donut (the farmer quadrants vs. all remaining quadrants) and are thus distinct from each other.

Column (5) in Table VII reports OLS estimates splitting the number of mills within 10 kms from the mill into farmer competition (mills in the farmer's quadrant) and mill competition (mills in the remaining quadrants). The estimates confirm a negative correlation between both measures of competition and the use of relational contracts as reported by the farmer.[36]

Column (8) explores the IV specification in which we separately instrument for farmer-competition and mill-competition. Columns (6) and (7) present the first-stages for farmer-level and mill-level competition respectively. Reassuringly, the two instruments are positively correlated with the corresponding measure of competition.

We find evidence that both mechanisms are at play. An additional competing mill in the farmer's quadrant reduces the relational contract index by 0.342 standard deviations. This is the effect of competition operating through both the direct temptation and the indirect profit mechanisms. An additional mill in other quadrants reduces the relational contract index by 0.223 standard deviations. This is the impact due to the indirect profit mechanism only. The difference between the two estimates, 0.119 standard deviations (p-value 0.09) isolates the direct temptation mechanism.

In industries with fixed costs and in which an additional entrant lowers output of incumbent firms (business stealing), average costs increase and entry is more desirable to the entrant than to the industry as a whole (Mankiw and Whinston (1986)). While this mechanism is similar to the indirect effect in our model, the two can be empirically distinguished. In our model, an additional entrant makes it harder to sustain

[36] Relative to the specification in Table II, there are two additional sources of measurement error in this specification. First, mills in other quadrants might also directly affect the farmer. Second, the process through which farmers are assigned to quadrants is noisy.

relational contracts for rivals, leading to a knock-on effect in which other relational contracts break down and aggregate amounts of cherries supplied to mills declines. In our context, instead, the model in Mankiw and Whinston (1986) would imply that an additional entrant raises the aggregate amount of cherries processed by mills (and increases the price paid to farmers).

This prediction is inconsistent with evidence that an additional competitor lowers the volume of cherries supplied by farmers to all mills (Table 6, Column 3). In the Online Appendix D we also show that (*1*) there is great abundance of coffee cherries to be processed and thus business stealing effects alone are unlikely to drive mill-level outcomes (see Online Appendix Figure D2); (*2*) our instrument for competition displays an inverted-U shaped relationship with the aggregate amount of cherries processed (see Panel B, Online Appendix Figure D3): as predicted by our model, past a certain level of entry, aggregate volumes processed decline with further entry.

# 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 key market failures and inform policy, particularly in developing economies.

This article 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 performance. Relational practices are, by definition, hard to codify and context-specific. While the practices we measure are relevant in our setting and in other agricultural value chains in developing countries, we hope to offer an example of the value to measure relational contracts 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: the design of adequate industrial policies needs to take into account informal arrangements and market institutions operating in specific contexts. 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.

The evidence suggests the possibility of excessive entry when contracts are hard to enforce. A direct policy recommendation, 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.[37] Such policy interventions, however, must be evaluated on a case-by-case basis since partial improvements in contract enforcement could undermine relationships and worsen market functioning (Baker et al. (1994)).

The industry in this article harbours (too) many unproductive firms − a rather typical portray of markets in developing economies. In such contexts, processes of consolidation (e.g., ownership changes through mergers and acquisitions) can potentially reduce inefficiencies but are often stifled by dysfunctional institutional environments. Indeed, in our context, such a process has started to emerge in recent years as more productive and better managed foreign exporters have acquired mills to integrate backwards (see Macchiavello and Morjaria (2020)).

[37]For example, in Costa Rica the Instituto del Cafe de Costa Rica (ICAFE) monitors the coffee value chain and enforces contracts between mills and farmers and between mills and exporters.

## References


Andrabi, T., J. Das, and I. Khwaja (2017) "Report Cards: The Impact of Providing School and Child Test Scores on Educational Markets," *American Economic Review*, Vol. 107, No. 6, pp. 1535–1563.

Antras, P. and C. Foley (2015) "Poultry in Motion: A Study of International Trade Finance Practices," *Journal of Political Economy*, Vol. 123, No. 4, pp. 809–852.

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

Baker, G., R. Gibbons, and K. J. Murphy (1994) "Subjective Performance Measures in Optimal Incentive Contract," *Quarterly Journal of Economics*, Vol. 109, No. 4, pp. 1125–1156.

——— (2002) "Relational Contracts and the Theory of the Firm," *Quarterly Journal of Economics*, Vol. 117, No. 1, pp. 39–84.

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

Bardhan, P. K. ed (1989) *The Economic Theory of Agrarian Institutions*, Oxford: Clarendon Press.

Bloom, N., E. Brynjolfsson, L. Foster, R. Jarmin, M. Patnaik, I. Saporta-Eksten, and J. Van Reenen (2017) "What Drives Differences in Management?" Working Paper 23300, NBER.

Bloom, N., C. Propper, S. Seiler, and J. Van Reenen (2015) "The Impact of Competition on Management Practices in Public Hospitals," *Review of Economic Studies*, Vol. 82, No. 2, pp. 457–489.

Bloom, N. and J. Van Reenen (2007) "Measuring and Explaining Management Practices across Firms and Countries," *Journal of Economic Perspectives*, Vol. 24, p. 1.

——— (2010) "Why do Management Practices Differ across Firms and Countries," *Annual Review of Economics*, Vol. 2, No. 1, pp. 105–137.

Blouin, A. and R. Macchiavello (2019) "Strategic Default in the International Coffee Market," *Quarterly Journal of Economics*, Vol. 134, No. 2, pp. 895–951.

Brynjolfsson, E. and P. Milgrom (2013) "Complementarity in Organizations," in Gibbons ed. *The Handbook of Organizational Economics*: Princeton University Press, pp. 680–731.

Burkart, M. and T. Ellingsen (2004) "In-Kind Finance: A Theory of Trade Credit," *American Economic Review*, Vol. 94, No. 3, pp. 569–590.

Cajal, J., R. Macchiavello, and G. Noguera (2020) "Buyers' Sourcing Strategies and Suppliers' Markups in Bangadeshi Garments," Working Paper, LSE.

Casaburi, L. and R. Macchiavello (2019) "Demand and Supply of Infrequent Payments as Commitment Device: Evidence from the Kenyan Dairy Industry," *American Economic Review*, Vol. 109, No. 2, pp. 523–555.

Casaburi, L. and T. Reed (2020) "Interlinked Transactions and Competition: Experimental Evidence from Cocoa Markets," Working Paper, Zurich.

Casaburi, L. and J. Willis (2018) "Time versus State in Insurance: Experimental Evidence from Contract Farming in Kenya," *American Economic Review*, Vol. 108, No. 12, pp. 3778–3813.

Conley, T. (1999) "GMM Estimation with Cross Sectional Dependence," *Journal of Econometrics*, Vol. 92, No. 1, pp. 1–45.

Djankov, S., R. La Porta, F. Lopez-de Silanes, and A Schleifer (2003) "Courts," *Quarterly Journal of Economics*, Vol. 118, No. 2, pp. 453–517.

Donaldson, D. and R. Hornbeck (2016) "Railroads and American Economic Growth: A 'Market-Access' Approach," *Quarterly Journal of Economics*, Vol. 131, No. 2, pp. 799–858.

Emran, S., D. Mookherjee, F. Shilpi, and M. Helal Uddin (2020) "Credit Rationing and Pass-Through in Supply Chains: Theory and Evidence from Bangladesh," Working Paper, Boston University.

Fisman, R. and M. Raturi (2004) "Does Competition Encourage Credit Provision? Evidence from African Trade Credit Relationships," *Review of Economic and Statistics*, Vol. 86, No. 1, pp. 345–352.

Ghani, T. and T. Reed (2020) “Relationships on the Rocks: Contract Evolution in a Market for Ice,” Working Paper, Washington University Olin Business School.

Gibbons, R. and R. Henderson (2012) “What Do Managers Do?: Exploring Persistent Performance Differences Among Seemingly Similar Enterprises,” in Gibbons ed. *The Handbook of Organizational Economics*: Princeton University Press, pp. 680–731.

Greif, A. (1993) “Contract Enforceability and Economic Institutions in the Early Trade: The Maghribi traders’ coalition,” *American Economic Review*, Vol. 83, No. 3, pp. 525–548.

Helper, S. and R. Henderson (2014) “Management Practices, Relational Contracts, and the Decline of General Motors,” *Journal of Economic Perspectives*, Vol. 28, No. 1, pp. 49–72.

Hsieh, C. and B. Olken (2014) “The Missing ‘Missing Middle’,” *Journal of Economic Perspective*, Vol. 28, No. 3, pp. 89–108.

Jensen, R. T. and N. H. Miller (2018) “Market Integration, Demand, and the Growth of Firms: Evidence from a Natural Experiment in India,” *American Economic Review*, Vol. 108, No. 12, pp. 3583–3625.

Konig, M. D., D. Rohner, M. Thoenig, and F. Zilibotti (2017) “Networks in Conflict: Theory and Evidence From the Great War of Africa,” *Econometrica*, Vol. 85, No. 4, pp. 1093–1132.

Macchiavello, R. (2010) “Development Uncorked: Reputation Acquisition in the New Market for Chilean Wines in the UK,” Discussion Paper 7698, CEPR.

Macchiavello, R. and J. Miquel-Florensa (2019) “Buyer-Driven Upgrading in GVCs: The Sustainable Quality Program in Colombia,” Working Paper, London School of Economics.

Macchiavello, R. and A. Morjaria (2015a) “Fully Washed Coffee Exports in Rwanda: Market Structure and Policy Implications,” *International Growth Centre, Policy Brief*.

——— (2015b) “The Value of Relationships: Evidence from a Supply Shock to Kenyan Rose Exports,” *American Economic Review*, Vol. 105, No. 9, pp. 2911–2945.

——— (2020) “Ownership Changes, Management and Efficiency: Evidence from Rwanda’s Coffee Industry,” Working Paper, Kellogg School of Management.

Mankiw, N. and M. D. Whinston (1986) "Free Entry and Social Inefficiency," *The RAND Journal of Economics*, Vol. 17, No. 1, pp. 48–58.

McMillan, J. and C. Woodruff (1999) "Interfirm Relationships and Informal Credit in Vietnam," *Quarterly Journal of Economics*, Vol. 114, pp. 1285–1320.

Minten, B., M. Dereje, E. Engida, and S. Tamru (2018) "Tracking the Quality Premium of Certified Coffee: Evidence from Ethiopia," *World Development*, Vol. 101, pp. 119–132.

Morjaria, A. and M. Sprott (2018) "Ugandan Arabica Coffee Value Chain Opportunities," *International Growth Centre, Policy Brief*.

Petersen, M. A. and R. G. Rajan (1995) "The Effect of Credit Market Competition on Lending Relationships," *Quarterly Journal of Economics*, Vol. 110, No. 2, pp. 407–443.

Rodrik, D. (2008) "Second-Best Institutions," *American Economic Review*, Vol. 98, No. 2, pp. 100–104.

Schilling, M. and T. McConnell (2004) "Model for Siting Coffee Washing in Rwanda," *Environmental System Research Institute, Inc*, Vol. 1697, Paper.

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*, Vol. 113, No. 3, pp. 582–625.

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

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

Stigler, G. (1956) "The Statistics of Monopoly and Merger," *Journal of Political Economy*, Vol. 64, No. 1, pp. 33–40.

Syverson, C. (2004) "Product Substitutability and Productivity Dispersion," *Review of Economics and Statistics*, Vol. 86, No. 2, pp. 534–550.

Table I: **SUMMARY STATISTICS**

| | (1) Mean | (2) $25^{th}$ Pct. | (3) Median | (4) $75^{th}$ Pct. | (5) Obs |
|---|---|---|---|---|---|
| Panel A: Mill Characteristics | | | | | |
| Mill age, years | 4.090 | 2 | 4 | 6 | 178 |
| Theoretical Capacity (tons of cherries) | 423.1 | 250 | 340.9 | 500 | 173 |
| Production (tons of parchment coffee) | 46.01 | 15 | 32 | 60 | 177 |
| Cherries Purchased (tons) | 294.8 | 102.4 | 199.9 | 400 | 174 |
| Seasonal employees | 35.13 | 16 | 30 | 50 | 171 |
| Cooperative status, dummy | 0.466 | 0 | 0 | 1 | 178 |
| Farmers in catchment area that sell to mill | 396.0 | 170 | 310 | 500 | 170 |
| NGO-supported mill, dummy | 0.264 | 0 | 0 | 1 | 178 |
| Total Unit Cost (RWF per kg) | 1793 | 1600 | 1800 | 1956 | 178 |
| Total Processing Unit Cost (RWF per kg) | 705.3 | 500.0 | 699.0 | 831.0 | 177 |
| Number of mills within 10km | 6.539 | 3 | 6 | 10 | 178 |
| Score within 5 km of mill | - | -0.826 | -0.276 | 0.714 | 177 |
| Score within 5-10 km of mill | - | -0.762 | -0.230 | 0.648 | 177 |
| Average Elevation (m) within 5km | 1625.8 | 1511 | 1630.4 | 1730.1 | 177 |
| Average Slope (°) within 5km | 10.93 | 8.859 | 10.87 | 12.87 | 177 |
| Average River density (m) within 5km | 320.5 | 205.5 | 319.5 | 423.2 | 177 |
| Average Tree density within 5km ('000) | 11.53 | 5.152 | 9.499 | 14.64 | 177 |
| Average Spring presence within 5km | 0.033 | 0.012 | 0.025 | 0.049 | 177 |
| Kilometers of road within 5km | 1.769 | 1.452 | 1.674 | 2.008 | 177 |
| Overall Quality Score | - | -0.473 | 0.150 | 0.745 | 159 |
| 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 |
| RC Index, mill outcomes | - | -0.894 | 0.252 | 0.252 | 177 |
| RC Index, overall | - | -0.502 | 0.114 | 0.453 | 175 |
| Panel B: Farmer Characteristics | | | | | |
| Farmer age, years | 46.44 | 36 | 47 | 56 | 875 |
| Female, dummy | 0.287 | 0 | 0 | 1 | 881 |
| Schooling, years | 5.339 | 4 | 6 | 7 | 879 |
| Distance to mill, km | 5.480 | 1.194 | 2.689 | 7.182 | 615 |
| Cooperative membership, dummy | 0.552 | 0 | 1 | 1 | 881 |
| Farmer's Trees | 975.5 | 250 | 500 | 1000 | 881 |
| Cherry price (RWF per kg) | 208.2 | 200 | 200 | 220 | 881 |
| Share sold as cherries (%) | 0.792 | 0.764 | 1 | 1 | 872 |
| Home process for saving, dummy | 0.232 | 0 | 0 | 0 | 881 |
| Job satisfaction index | - | -0.457 | 0.026 | 0.499 | 868 |
| Number of other mills in own quadrant | 1.506 | 0 | 1 | 2 | 615 |
| 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 | 0 | 1 | 1 | 877 |
| RC Index, farmer outcomes | - | -0.659 | 0.413 | 0.413 | 881 |

*Note:* Mill characteristics are obtained from the survey of mills and author's GIS dataset. Farmer characteristics are obtained from a survey of 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. Relational contract index measures, referred to as relational practices in the text are dummy variables: Given inputs to farmers, Has made a second payment in the past and Provides help/loans to farmers are responses from mill managers and Received input from mill, Expects to receive a second payment and Expects to receive help/loan are responses from farmer surveys. Competition is defined as the number of mills within 10km. Means of standardized variables are denoted by "-" denote standardized indices (z-score's).

## Table II: **COMPETITION AND RELATIONAL CONTRACTS**

| Dependent Variable | (1)<br>RC Index<br>(z-score) | (2)<br>Competition | (3)<br>RC Index<br>(z-score) | (4)<br>RC Index<br>(z-score) |
|---|---|---|---|---|
| Competition | -0.116<br>(0.025)***<br><0.025>***<br>[0.026]*** | | | -0.283<br>(0.095)***<br><0.075>***<br>[0.060]*** |
| Score within 5-10 km of mill | | 1.610<br>(0.329)***<br><0.313>***<br>[0.419]*** | -0.455<br>(0.105)***<br><0.107>***<br>[0.063]*** | |
| Score within 5 km of mill | YES | YES | YES | YES |
| Geographic controls | YES | YES | YES | YES |
| Mill controls | YES | YES | YES | YES |
| Adjusted $R^2$ | 0.27 | 0.70 | 0.25 | 0.10 |
| Observations | 175 | 177 | 175 | 175 |
| Model | OLS | First stage | Reduced | IV |

*Note:* Standard errors are denoted as follows: (Bootstrap in which mills are re-sampled 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]. Stars * * * (**) [*] indicate significance at the 0.01 (0.05) [0.1] level. The RC index is a z-score of the three aggregate indices: pre-harvest, harvest and post-harvest indices with equal weighting of farmer and mill responses. 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 promised second payments post-harvest. Post-harvest z-score is constructed from farmer- and mill manager-based indicators loans or help provided after the harvest. Competition is measured as the number of mills within a 10 km radius, and is instrumented with the engineering model suitability score in locations 5 km to 10 km away from the mill (referred to in the table as "Score within 5-10 km of mill". For ease of comparison between the OLS and IV estimates, Column 1 already includes the average suitability score within 5 km as a control (referred to in the table as "Score within 5 km of mill"). The average suitability scores from the engineering model are all z-scores. Mill controls include whether the mill is NGO-supported, cooperative status, mill age, mill age squared, and mill coordinates. Geographic controls include average engineering suitability score, average spring presence, road density, tree density, rivers, flexible control for coffee suitability, elevation, and slope, all within 5 km of the mill. Coffee suitability is from the FAO's Global Agro-Ecological Zones (FAO-GAEZ) dataset. Estimates for crop suitability are available for various input levels. To match conditions for Rwanda, the data chosen was for low-input and rain-fed conditions. The resolution is at the 5 arc-minute level, at the equator that is almost a resolution of 9 km × 9 km, see http://www.fao.org/nr/gaez/about-data-portal/agricultural-suitability-and-potential-yields/en/, accessed August 2014.

Table III: **UNPACKING COMPLEMENATARY RELATIONAL PRACTICES**

| Dependent Variable | (1) Received input from mill | (2) Given inputs to farmers | (3) RC Pre-Harvest Z-score | (4) Expects to receive a second payment | (5) Has made a second payment in the past | (6) RC Harvest Z-score | (7) Expects to receive help/loan | (8) Provides help/loans to farmers | (9) RC Post-Harvest Z-score | (10) RC Index | (11) RC Index, farmer outcomes | (12) RC Index, mill outcomes | (13) Placebo: Short-Term Credit |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Panel A: IV** | | | | | | | | | | | | | |
| Competition | -0.064*** | -0.085** | -0.220** | -0.063*** | -0.077** | -0.203*** | -0.066*** | -0.026 | -0.180* | -0.283*** | -0.237*** | -0.215** | 0.017 |
| | (0.014) | (0.036) | (0.112) | (0.017) | (0.035) | (0.074) | (0.020) | (0.034) | (0.099) | (0.098) | (0.045) | (0.094) | (0.082) |
| **Panel B: OLS** | | | | | | | | | | | | | |
| Competition | -0.011** | -0.030** | -0.062*** | -0.038*** | -0.041*** | -0.121*** | -0.021*** | -0.015 | -0.065** | -0.116*** | -0.086*** | -0.102*** | -0.041 |
| | (0.005) | (0.012) | (0.022) | (0.006) | (0.013) | (0.038) | (0.007) | (0.014) | (0.033) | (0.029) | (0.016) | (0.029) | (0.033) |
| 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 $R^2$ | 0.05 | 0.13 | 0.11 | 0.15 | 0.16 | 0.21 | 0.03 | -0.01 | 0.00 | 0.10 | 0.07 | 0.12 | 0.04 |
| Observations | 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. RC Pre-harvest z-score is constructed based on farmer- and mill manager-based indicators of mill-provided inputs. RC Harvest z-score is constructed from farmer- and mill manager-based indicators of promised second payments post-harvest. RC Post-harvest z-score is constructed from farmer- and mill manager-based indicators of loans or help provided after the harvest. The RC Index (column 10) is an aggregate of these three indices with responses from the farmer and the mill manager equally weighted. Aggregate relational practices from the farmers' perspective and mill managers' perspective are reported in column 11 and column 12, respectively. Column 13 captures if there was any short-term credit provided to the farmer by the mill. Mill controls include NGO-supported, cooperative status, mill age, mill age squared and mill coordinates. Geographic controls includes average engineering score, average spring presence, road density, tree density, rivers, flexible control of FAO-GAEZ coffee suitability, elevation and slope, all within 5 km of the mill. Farmer controls include farmer age, education, gender, schooling, distance to mill, cognitive test (z-score) and cooperative membership. Farmer responses are from the 2012 farmer survey. Competition is measured as the number of mills within a 10 km radius and is instrumented with the engineering model suitability score in locations 5 km to 10 km away from the mill. Adjusted $R^2$ is provided for Panel A (IV). For additional variable definitions refer to notes in Table II.

## Table IV: **COMPETITION AND MILL OUTCOMES**

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
|---|---|---|---|---|---|---|---|---|---|---|
| Dependent Variable | Total Unit Cost (RWF,ln) | Cherry Price (RWF,ln) | Home-processed Parchment Price (RWF,ln) | Conversion Ratio (ln) | Total Processing Unit Cost (RWF,ln) | Cherries Purchased (Tons) | Weeks Mill Processed | Days with too many workers | Days with too few workers | Unit Labour Costs (RWF,ln) |
| **Panel A: IV** | | | | | | | | | | |
| Competition | 0.046* | 0.013 | -0.007 | -0.007 | 0.071** | -4.984* | -0.202 | 0.227** | 0.095* | 0.107* |
| | (0.026) | (0.013) | (0.011) | (0.016) | (0.031) | (2.877) | (0.322) | (0.099) | (0.055) | (0.058) |
| **Panel B: OLS** | | | | | | | | | | |
| Competition | 0.010* | 0.002 | -0.001 | -0.011*** | 0.017 | -0.733 | 0.090 | 0.107*** | 0.044* | 0.068*** |
| | (0.005) | (0.004) | (0.003) | (0.004) | (0.011) | (0.591) | (0.142) | (0.032) | (0.026) | (0.019) |
| 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 $R^2$ | . | 0.24 | 0.28 | -0.02 | -0.08 | 0.24 | 0.08 | 0.13 | 0.01 | 0.16 |
| Observations | 177 | 163 | 176 | 145 | 176 | 169 | 160 | 173 | 174 | 176 |
| Dep.var. mean | 7.48 | 5.34 | 6.64 | 4.17 | 6.51 | 27.17 | 15.11 | 2.20 | 1.29 | 5.44 |

*Note:* Bootstrapped standard errors in parentheses. $***$ ($**$) [$*$] indicates significance at the 0.01 (0.05) [0.1] level. Column 1 is the total unit cost of producing 1 kilogram of the output (parchment). Mill managers' response to prices paid to farmers for cherries (ln) and local home processed parchment price (ln) are reported in columns 2 and 3 respectively. Column 4 is the physical efficiency ratio between the input (cherry) and output (parchment). Column 5 reports the processing cost (excluding cost of cherry) from the unit cost of processing reported in column 1. Column 6 is the total cherries purchased by the mill as reported to the coffee board in the 2012 season, mill capacity is controlled for in this specification. Column 7 is the number of weeks the mill is operational in the 2012 season. In columns 8 and 9 the dependent variables, "Days with too many workers" and "Days with too few workers" are all dummy variables reported by the mill manager. Competition is measured as the number of mills within 10 km, and is instrumented with the engineering model suitability score in locations 5 km to 10 km away from the mill. Mill controls include NGO-supported, cooperative status, mill age, mill age squared, and mill coordinates. Engineering controls and Geographical controls include average engineering score, average spring presence, road density, tree density, rivers, flexible control of FAO-GAEZ coffee suitability, elevation and slope, all within 5 km of the mill. Adjusted $R^2$ is provided for Panel A (IV). Responses are provided by the mill manager. For additional variable definitions refer to notes in Table II.

Table V: **COMPETITION AND QUALITY OF MILL OUTPUT**

| Dependent Variable | (1) Overall Quality Score | (2) Farmer Controlled Quality | (3) Mill Controlled Quality | (4) Plant Genetic Properties |
|---|---|---|---|---|
| **Panel A: IV** | | | | |
| Competition | -0.146* | -0.172** | -0.034 | 0.018 |
| | (0.087) | (0.078) | (0.088) | (0.054) |
| **Panel B: OLS** | | | | |
| Competition | -0.039 | -0.051 | -0.043 | 0.024 |
| | (0.033) | (0.034) | (0.035) | (0.026) |
| 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 $R^2$ | -0.07 | 0.04 | 0.19 | -0.01 |
| Observations | 158 | 155 | 156 | 157 |

*Note:* Bootstrapped standard errors in parentheses. $***$ ($**$) [$*$] indicates significance at the 0.01 (0.05) [0.1] level. Column 1 is the overall quality score in constructed from the farmer and mill quality indices with equal weighting to all indicators, plus an indicator of ideal conversion ratio, an indicator of specialty status, and standardized cupping points. Column 2 is the farmer-controlled quality outcome as a standardized index of an indicator of large beans and of severe insect damage. Column 3 is the mill-controlled quality index constructed from an indicator of high moisture, of floaters, and of broken beans. All components of indices are re-scaled so that higher values indicate higher quality. Mill controls include NGO-supported, cooperative status, mill age, mill age squared and mill coordinates. Engineering controls and Geographical controls include average engineering suitability score, average spring presence, road density, tree density, rivers, flexible control of FAO-GAEZ coffee suitability, elevation, and slope, all within 5 km of the mill. Competition measure is the number of mills within a 10 km radius, and is instrumented with the engineering model suitability score in locations 5 km to 10 km away from the mill. Adjusted $R^2$ is provided for Panel A (IV). For additional variable definitions refer to notes in Table II.

## Table VI: **COMPETITION AND FARMER OUTCOMES**

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) | (13) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Dependent Variable | Cherry price (RWF ,ln) | Home-processed Parch-ment Price (RWF,ln) | Share sold as cherries (%) | Home process for saving, dummy | Self financed inputs | Yield (ln) | Input usage (RWF,ln) | Farmer's trees (ln) | Farmer revenues (RWF,ln) | Job satis-faction index | Pay is good | Freedom to decide how to do my job | Work is stressful |
| **Panel A: IV** | | | | | | | | | | | | | |
| Competition | 0.012*** | -0.004 | -0.075*** | 0.059*** | 0.034** | -0.004 | -0.024 | 0.057 | -0.081* | -0.059** | -0.065** | -0.047** | 0.076*** |
| | (0.003) | (0.004) | (0.016) | (0.019) | (0.017) | (0.031) | (0.064) | (0.048) | (0.045) | (0.023) | (0.032) | (0.024) | (0.029) |
| **Panel B: OLS** | | | | | | | | | | | | | |
| Competition | 0.004*** | 0.007*** | 0.003 | 0.007 | 0.025*** | 0.039*** | -0.016 | 0.017 | 0.083*** | -0.035*** | -0.030*** | -0.035*** | -0.005 |
| | (0.001) | (0.001) | (0.004) | (0.007) | (0.007) | (0.011) | (0.024) | (0.017) | (0.016) | (0.007) | (0.010) | (0.007) | (0.011) |
| Score within 5 km of mill | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Engineering controls | 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 | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Adjusted $R^2$ | 0.26 | 0.32 | . | 0.00 | 0.08 | 0.14 | 0.06 | 0.14 | 0.36 | 0.12 | 0.11 | 0.07 | -0.02 |
| Observations | 869 | 869 | 860 | 869 | 869 | 865 | 472 | 869 | 848 | 856 | 855 | 856 | 855 |
| Dep. var. mean | 5.33 | 6.63 | 0.79 | 0.23 | 0.35 | 0.38 | 3.30 | 6.22 | 4.84 | - | 2.96 | 3.61 | 3.48 |

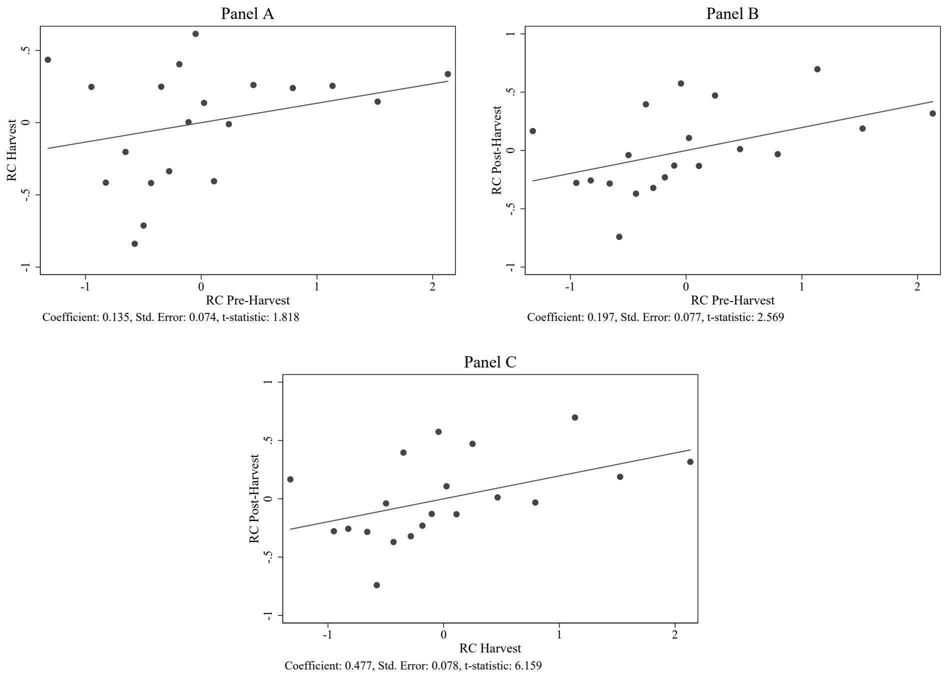
*Note:* Bootstrapped standard errors in parentheses. $***$ ($**$) [$*$] indicates significance at the 0.01 (0.05) [0.1] level. Column 1 reports farmers' cherry price and column 2 reports local home processed parchment price. Column 3 reports share of farmers' coffee output sold as cherries to mills. Column 4 is the farmers' response to an open ended question on why they would home process coffee, with a dummy variable created for the responses that states saving as the main motive. Column 5 is a dummy variable indicating if the farmer self-financed input purchases for their plot. Column 6 is the yield (in ln) and column 7 is the amount of self-financed inputs per kg of output (cherries). Column 8 is the farmer's tree holdings (ln) and column 9 are the farmer revenues (sums of revenues received from coffee sold as cherries and as home processed parchment). Column 10,"Job Satisfaction" Index, is a z-score 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". Columns 11 to 13 are some of the individual components of the Job Satisfaction Index. Mill controls include NGO-supported, cooperative status, mill age, mill age squared, and mill coordinates. Engineering controls and Geographical controls include average engineering score, average spring presence, road density, tree density, rivers, flexible control of FAO-GAEZ coffee suitability, elevation and slope, all within 5 km of the mill. Competition is measured as the number of mills within 10 km, and is instrumented with the engineering model score in locations 5 km to 10 km away from the mill. Farmer controls include farmer age, education, gender, schooling, distance to mill, cognitive score, cooperative membership and log number of coffee trees (except in column 7). All farmer outcomes/responses are from the 2012 farmer survey. Adjusted $R^2$ is provided for Panel A (IV).

Table VII: **RELATIONAL CONTRACTING AND COMPETITION - FARMER LEVEL**

| | (1) Baseline | (2) Reduced sample | (3) Two nearest mills | (4) Second nearest mill only | (5) RC Index, farmer outcomes | (6) Farmer Competition | (7) Mill Competition | (8) RC Index, farmer outcomes |
|---|---|---|---|---|---|---|---|---|
| Competition | -0.237*** | -0.213*** | -0.205*** | -0.209*** | | | | |
| | (0.039) | (0.064) | (0.057) | (0.046) | | | | |
| Farmer Mill-Access | | | -0.042 | -0.083* | | | | |
| | | | (0.027) | (0.044) | | | | |
| Farmer-level Competition | | | | | -0.055** | | | -0.342*** |
| | | | | | (0.026) | | | (0.111) |
| Mill-level Competition | | | | | -0.103*** | | | -0.223** |
| | | | | | (0.017) | | | (0.087) |
| Engineering Score in own quadrant | | | | | | 0.870*** | -0.579*** | |
| | | | | | | (0.098) | (0.121) | |
| Engineering Score in other quadrants | | | | | | -0.274*** | 1.417*** | |
| | | | | | | (0.098) | (0.198) | |
| Equality of Coefficients, p-value | | | | | | | | 0.039 |
| Competition Definition | | | | | # 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 | YES | YES |
| District fixed effects | | | | | YES | YES | YES | YES |
| Farmer controls | | | | | YES | YES | YES | YES |
| Quadrant fixed effects | | | | | YES | YES | YES | YES |
| OLS/IV | | | | | OLS | OLS | OLS | IV |
| Adjusted $R^2$ | 0.07 | 0.11 | 0.13 | 0.12 | 0.18 | 0.47 | 0.64 | 0.00 |
| Observations | 869 | 606 | 606 | 606 | 606 | 606 | 606 | 606 |

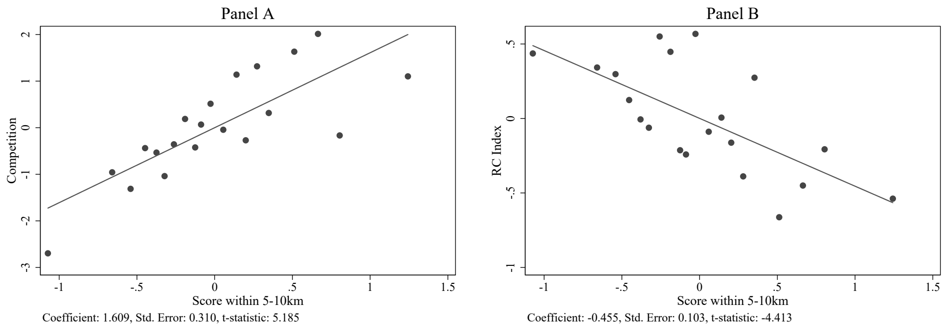
*Note:* Bootstrapped standard errors in parentheses. $***$ ($**$) [$*$] indicates significance at the 0.01 (0.05) [0.1] level. The dependent variable across all columns is the Relational Contract Index for farmer outcomes. Column 3 includes a "Farmer Mill-Access" this is a measure of a farmers' access to a mill, analogous to Donaldson and Hornbeck (2016); the measure is constructed as the inverse sum of the distances to the nearest and second nearest competing mills to the farmer with no restriction on the radius. Column 4 the "Farmer Mill-Access" measure is now the inverse of the distance to the second nearest competitor with no restriction on the radius. Farmer controls include distance to mill, farmer age, education, gender, schooling, cognitive score, cooperative membership, and holdings of coffee trees (ln). Mill controls include NGO-supported, cooperative status, mill age, mill age squared and mill coordinates. Additional geographical controls include the average engineering suitability score, average spring presence, road density, tree density, rivers, flexible control of FAO-GAEZ coffee suitability, elevation, slope, all within 5 kms of the mill. 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 Relational Contract Index measure 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 we only have exact village location information for 70% of our surveyed farmers in 2012. *P-value* is reported for the test that farmer and mill effects are equal.

Figure I: **RELATIONAL CONTRACT PRACTICES**




*Note:* Binned scatter plot of mill-level regressions. All regressions control for mill characteristics (NGO-support, cooperative status, mill age, mill age squared and mill coordinates). Controls also include average engineering suitability score, average spring presence, road density, tree density, rivers, flexible control of FAO-GAEZ coffee suitability, elevation and slope, all within 5 km of the mill. RC Pre-harvest (z-score) is constructed based on farmer- and mill manager- based indicators of mill-provided inputs. RC Harvest (z-score) is constructed from farmer- and mill manager- based indicators of trade credit and second payments. RC 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. In all RC index z-score's, farmer and mill manager responses are equally weighted.

Figure II: **INSTRUMENTAL VARIABLE - FIRST STAGE AND REDUCED FORM**




*Note:* Binned scatter plot of mill-level regressions. All regressions control for mill characteristics (NGO-support, cooperative status, mill age, mill age squared and mill coordinates). Controls also include average engineering suitability score, average spring presence, road density, tree density, rivers, flexible control of FAO-GAEZ coffee suitability, elevation and slope, all within 5 km of the mill. The RC Index is an aggregate of farmer- and mill manager- based indicators of mill-provided inputs, second payments, and post-harvest loans. Farmer and mill manager responses are equally weighted. Competition is measured as the number of mills within 10 km.

**NOT FOR PUBLICATION**

**Online Appendix**

# Competition and Relational Contracts in the Rwanda Coffee Chain

***by*** **Rocco Macchiavello and Ameet Morjaria**

**November 2020**

## Table of Contents

# A Theory

*Incentive Compatibility Constraints*: This section spells out all the incentive compatibility constraints associated with the relational contract. We derive conditions under which the following actions occur in each period of a stationary relational contract: $(i)$ at pre-harvest either $\mathbf{I}_{k_i} = 1$ or $\mathbf{I}_{k_i} = 0$; $(ii)$ the farmer sells to the mill, $x_i = 1$; and $(iii)$ the mill makes payments $P_{i,1}$ and $P_{i,2}$ at harvest and post-harvest.

Several incentive compatibility constraints must be met for such relational contract to be self-enforcing. First, if the relational contract specifies $\mathbf{I}_{k_i} = 1$, the mill must be willing to incur cost $k$ pre-harvest rather than not doing so and ceasing all interactions with the farmer. This gives the incentive constraint:

$$((v-c)(1+\pi)q-P_{i,1}-P_{i,2}-k)+\frac{\delta}{1-\delta}((v-c)(1+\pi)q-P_{i,1}-P_{i,2}-k) \geq 0+\frac{\delta}{1-\delta}\times 0. \quad (9)$$

Second, at harvest, the farmer must prefer to sell to the mill rather than taking her outside option and then losing access to the mill in the future.[38] Note that the farmer's per-period payoff in the relational contract is given by $u(c_{i,1}, c_{i,2}) = \min\{P_{i,1}, P_{i,2}\}$. If the farmer side-sells she gets $(1 + \mathbf{I}_{k_i}\pi)q \times u(\rho, C_i)$ this season and her outside option $q \times u(\rho, C_i)$ forever after. This gives the no side-selling incentive constraint:

$$min\{P_{i,1}, P_{i,2}\} + \frac{\delta}{1-\delta} min\{P_{i,1}, P_{i,2}\} \geq (1 + \mathbf{I}_{k_i}\pi)qu(\rho, C_i) + \frac{\delta}{1-\delta} qu(\rho, C_i) \quad (10)$$

Third, at harvest the mill must be willing to pay $P_{i,1}$. If the mill does not pay $P_{i,1}$, we assume the farmer does not deliver any coffee to the mill. The incentive constraint is given by:

$$((v-c)(1+\mathbf{I}_{k_i}\pi)q-P_{i,1}-P_{i,2})+\frac{\delta}{1-\delta}((v-c)(1+\mathbf{I}_{k_i}\pi)q-P_{i,1}-P_{i,2}-\mathbf{I}_{k_i}k) \geq 0. \quad (11)$$

Finally, post-harvest, the mill must prefer to pay the second payment $P_{i,2}$ and continue the relationship rather than defaulting and obtaining her outside option equal to zero from then onward. The incentive constraint is given by:

$$-P_{i,2} + \frac{\delta}{1-\delta}((v-c)(1+\mathbf{I}_{k_i}\pi)q - P_{i,1} - P_{i,2} - \mathbf{I}_{k_i}k) \geq \frac{\delta}{1-\delta} \times 0. \quad (12)$$

[38] Note that we abstract from the possibility that the farmer might divert the inputs provided by the mill for private consumption. If that was the case, the associated incentive constraint could become binding and the no side-selling constraint slack. The farmer's continuation value following a deviation would still depend on competition and, therefore, the main insights of our analysis would be preserved.

Note that we treat harvest-time transaction as a spot transaction, in which $P_{i,1}1$ and cherries are exchanged simultaneously.[39]

*Which incentive constraints bind?* Since the relational contract maximizes the mill profits, it must be that $P_{i,1} = P_{i,2}$ and that (10) is binding. This implies

$$P_{i,1} = P_{i,2} = ((1-\delta)(1+\mathbf{I}_{k_i}\pi)+\delta) \times qu(\rho, C_i) \quad (13)$$

Furthermore, note that (12) implies both (9) and (11). Substituting (13) into (12) we obtain the necessary condition under which a self-enforcing relational contract exists. The condition is given by:

$$\frac{\delta}{1-\delta}((v-c)(1+\mathbf{I}_{k_i}\pi)q - \mathbf{I}_{k_i}k) \geq u(\rho, C_i)q(1+(1-\delta)\mathbf{I}_{k_i}\pi) \quad (14)$$

This condition states that the net present value of the per period rents generated by selling cherries to the mill (given by $(v-c)(1+\mathbf{I}_{k_i}\pi)q - \mathbf{I}_{k_i}k)$) ought to be larger than the aggregate temptation to deviate (which is equal to the second payment $P_{i,2}$).

**Proof of Proposition 1 (Direct Effect of Competition)**: Statement 1) follows from the observation that the left-hand side of (14) is increasing in $\delta$ while the right-hand side is decreasing in $\delta$. Statement 2) follows from the observation that the right hand side of (14) is increasing in $C_i$. Finally, denote with $u^{'}_{\mathbf{I}_k}$ the level of the outside option $u(\rho, C_i)$ such that condition (14) holds with equality. Statement 3) is equivalent to the condition $u^{'}_0 > u^{'}_1$. ||

[39]This is not always literally what happens in the field, but it is not an essential feature for the mechanisms we are trying to capture. The assumption is in line with our empirical focus on relational practices involving exchanges of promises over several weeks as well as with farmers' reports that they are concerned about mills defaulting on second payments after harvest, rather than being held-up on cash payments during harvest. The assumption can however be relaxed without altering the main insights.

# B Additional Tables

## List of Tables

Table B1: National Coffee Census 2009 and Farmer Survey 2012

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Dependent Variable | Household size | Age (ln) | Number of Trees (ln) | Distance to Kigali (km) | Distance to Sector capital (km) | Distance to market (km) | Distance to nearest mill (km) | Elevation (m) | Slope (°) | FAO Coffee Suit-ability | River density (m) | Roads (km) |
| Farmer in Survey | 0.023 | -0.014 | 1.545*** | -0.142 | -0.107 | 0.086 | -0.533*** | -6.252 | -0.045 | 1.739 | -8.982 | -0.030 |
| | (0.014) | (0.015) | (0.061) | (0.340) | (0.184) | (0.133) | (0.066) | (6.417) | (0.145) | (1.216) | (23.867) | (0.046) |
| Observations | 244,927 | 222,516 | 244,612 | 243,313 | 230,841 | 244,214 | 244,927 | 244,927 | 244,927 | 244,927 | 244,927 | 244,927 |
| Dep.var. mean | 0.87 | 3.80 | 4.67 | 113.67 | 6.71 | 4.98 | 2.30 | 1655.16 | 11.89 | 498.47 | 295.68 | 1.86 |

*Note:* Standard errors are clustered at the sector-level. $***$ $(**)$ $[*]$ indicates significance at the 0.01 (0.05) [0.1] level. Dependent variables in Columns 1 to 3 are from the National Coffee Census of all coffee farmers in 2012. Column 1 is household size, defined as a dummy taking a value of 1, if the household has more than two members, zero otherwise. Column 2 is the farmers' age (ln) and column 3 is the farmers tree holding size (ln). Columns 4 to 12 are constructed from the GIS dataset of geographical variables. "Farmer in Survey" is a dummy variable that takes a value of 1 if our 2012 surveyed farmer is also in the 2009 National Coffee Census, zero otherwise. We restrict our sample to all farmers whose village centroids are within a 5 km radius of a mill. Sector fixed effects are included in all columns.

Table B2: Engineering Model for Optimal Mill Placement

| | (1) Mill Entry | (2) Mill Entry | (3) Mill Entry | (4) Mill Entry |
|---|---|---|---|---|
| Spring within grid box | 0.277 | | 0.279 | -2.703*** |
| | (0.172) | | (0.173) | (0.342) |
| Untarred Local Road within grid box | | 0.516*** | 0.516*** | 0.499*** |
| | | (0.169) | (0.169) | (0.169) |
| Spring × Untarred Local Road | | | | 2.997*** |
| | | | | (0.366) |
| Geographic Controls: | | | | |
| Polynomials | YES | YES | YES | YES |
| Interactions | YES | YES | YES | YES |
| pseudo-$R^2$ | 0.15 | 0.15 | 0.15 | 0.15 |
| Observations | 13759 | 13759 | 13759 | 13759 |

*Note:* Standard errors are clustered by sector. $***$ ($**$) [$*$] indicates significance at the 0.01 (0.05) [0.1] level. Estimation method is a probit model. Observations denote the number of grids (1 $km^2$) on the map of Rwanda. The dependent variable, Mill Entry is a dummy taking a value of 1 if a mill is present in the grid and zero otherwise. All regressions control for linear, quadratic, and cubic terms of elevation, slope, flexible control for FAO-GAEZ coffee suitability, rivers, size of administrative sector, coffee tree density, latitude and longitudinal coordinates of grid, as well as the interactions between each of these variables (indicated by "Interaction" in the Geographic controls). Coffee suitability is from the FAO's Global Agro-Ecological Zones (FAO-GAEZ) dataset. Estimates for crop suitability are available for various input levels. To match conditions for Rwanda, data chosen was for low-input and rain-fed conditions. The resolution is at the 5 arc-minute level, at the equator that is almost a resolution of 9 km × 9 km, see http://www.fao.org/nr/gaez/about-data-portal/agricultural-suitability-and-potential-yields/en/, accessed August 2014.

Table B3: Robustness to IV: Exclusion Restriction and Bad Controls

| | (1) Baseline | (2) No roads in IV | (3) No roads | (4) Baseline FAO Coffee | (5) No trees in IV | (6) No trees | (7) Trees 1999 in IV | (8) Trees 1999 | (9) No roads, no trees in IV | (10) OLS |
|---|---|---|---|---|---|---|---|---|---|---|
| **Panel A: IV** | | | | | | | | | | |
| Competition | -0.283*** | -0.290*** | -0.308*** | -0.279*** | -0.288*** | -0.182*** | -0.224*** | -0.225** | -0.295*** | -0.308*** |
| | (0.072) | (0.076) | (0.075) | (0.068) | (0.073) | (0.069) | (0.075) | (0.101) | (0.077) | (0.119) |
| **Panel B: OLS, First stage** | | | | | | | | | | |
| Score within 5-10 km of mill | 1.610*** | 1.538*** | 1.550*** | 1.634*** | 1.616*** | 1.495*** | 1.486*** | 1.234*** | 1.549*** | 1.149*** |
| | (0.312) | (0.323) | (0.317) | (0.285) | (0.311) | (0.337) | (0.283) | (0.311) | (0.323) | (0.323) |
| Score within 5 km of mill | 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 |
| Centered $R^2$ | 0.19 | 0.18 | 0.10 | 0.20 | 0.18 | 0.23 | 0.27 | 0.21 | 0.17 | 0.14 |
| Observations | 175 | 175 | 175 | 175 | 175 | 175 | 175 | 175 | 175 | 175 |
| Instrument Score, 5-10km | Trees 2009 | No roads | No roads | Trees 2009 | No trees | No trees | Trees 1999 | Trees 1999 | No roads, No trees | Trees 2009 |
| Entry Model | 30k trees | 30k trees | 30k trees | FAO Coffee | FAO Coffee | FAO Coffee | FAO Coffee | FAO Coffee | FAO Coffee | 30k trees |
| Catchment Area Score, 0-5km | Trees 2009 | Yes roads | No roads | Trees 2009 | Trees 2009 | No trees | Trees 2009 | Trees 1999 | Trees 2009 | Trees 2009 |
| Controls in Panel A and B | Trees 2009 | Yes roads | No roads | Trees 2009 | Trees 2009 | No trees | Trees 2009 | Trees 1999 | Trees 2009 | Trees 2009 |

*Note:* Standard errors adjusted for arbitrary spatial clustering using the *acreg* package written by Konig et al. (2017). $***$ ($**$) [$*$] indicates significance at the 0.01 (0.05) [0.1] level. The dependent variable across all the specification is the relational contract index (z-score). Column 1 is the baseline specification using the engineering model (Table B2, column 4). "Instrument Score, 5-10km" refers to the instrumental variable, the average suitability z-score in the donut area. "Entry Model" refers to the engineering model, specifically that mill placement should take place in sectors with at least 30,000 coffee trees. "Catchment Area Score, 0-5km" refers to the average suitability z-score in the catchment area of the mill. This robustness table drops various criterion of the engineering model (Table B2 and geographic controls in Panel A and B in comparison to the baseline specification (column 1). In column 2, the presence of roads is removed in the engineering model when constructing the instrument; that is, there are no roads in the construction of the suitability score in the donut area. In column 3, in addition to removing roads in the instrument (column 2), roads are also removed from the engineering model when constructing the suitability score for the catchment area. Further, roads are also excluded from geographical controls. In columns 1 to 3, coffee trees are extracted from the 2009 National Coffee Census. Each sector must have at least 30,000 coffee trees in order for the 1 $km^2$ grid boxes to have a positive score. Columns 4 to 9 systematically change how the presence of coffee trees enters the engineering model criterion, construction of the suitability score for the instrument and the catchment area as well as geographical controls. In column 4, we replace the 30,000 coffee tree restriction with the FAO-GAEZ coffee suitability measure (referred to as "FAO Coffee" in the table) while the rest of the specification remains the same as our baseline. In column 5, in addition to using the FAO-GAEZ coffee suitability (in lieu of actual coffee trees) as the criterion, trees are excluded in constructing the instrument (i.e. suitability score in the donut area). In column 6, in addition to the column 5 modification we also exclude tree density from geographical controls, leaving only the FAO-GAEZ suitability in the engineering criteria. Actual coffee trees are no longer included in the instrument; similarly, the catchment score is not included in the geographical controls. In column 7, actual coffee trees are used again but this time extracted from a period before the industry took off, the 1999 National Coffee Census. Trees from the 1999 census are used to construct the instrument (i.e. suitability in the donut area). In the entry model criterion we use the FAO-GAEZ coffee suitability. In the catchment suitability score and geographical controls actual trees from the 2009 National Coffee Census are used. In column 8, the engineering model estimation uses FAO-GAEZ coffee suitability instead of the 30,000 coffee trees criterion and coffee trees are derived from the 1999 National Coffee Census in the instrument, catchment suitability score and geographic controls. In column 9, we run a similar specification to column 5 but we exclude both roads and trees to construct the instrument. In column 10, we run the same specification as our baseline in column 1 except the engineering model is an OLS rather than a probit model; negative predictions censored to 0 and predictions above the $95^{th}$ percentile have been winsorized. Centered $R^2$ is provided for Panel A (IV).

Table B4: Farmer Characteristics, 2012 Survey & 2009 National Coffee Census

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Dependent Variable | Farmer Age (years, survey) | Gender | Schooling (years, survey) | Cognitive test (z-score, survey) | Household size (census) | Farmer Age (years, census) |
| **Panel A: IV** | | | | | | |
| Score within 5-10 km of mill | -0.204 | -0.033 | -0.044 | -0.087 | 0.009 | 0.331 |
| | (0.861) | (0.021) | (0.135) | (0.061) | (0.006) | (0.239) |
| **Panel B: Competition** | | | | | | |
| Competition | -0.670*** | 0.003 | 0.002 | -0.016 | -0.001 | -0.133* |
| | (0.218) | (0.007) | (0.039) | (0.011) | (0.001) | (0.068) |
| Engineering controls | YES | YES | YES | YES | YES | YES |
| Observations | 870 | 876 | 874 | 876 | 243356 | 221030 |

*Note:* Standard errors are clustered at the mill-level. $***$ ($**$) [$*$] indicates significance at the 0.01 (0.05) [0.1] level. Farmers are assigned to their nearest mill and outcomes in columns (1) to (4) are from the 2012 farmer survey. Farmer outcomes in columns (5) and (6) are from the 2009 National Coffee Census of all coffee farmers. Panel A shows the raw correlation between farmer outcomes and our competition instrument (Score within 5-10 km of the mill), controlling for the engineering suitability score within the catchment area (5 km). Panel B shows the raw correlation between farmer outcomes and competition (number of mills in the 10 km radius), controlling for the engineering suitability score within the catchment area (5 km). No other controls are included in the columns.

Table B5: Farmer 2019 Survey and Outside Economic Opportunities

| Dependent Variable | (1) Non-coffee Income (%) | (2) Non-coffee Income | (3) Worked for someone | (4) Days worked for someone (ln) | (5) Wage rate (RWF,ln) | (6) Total payment received (RWF,ln) | (7) Pay if hire someone (RWF,ln) | (8) Sell milk | (9) Milk price (RWF,ln) | (10) Milk revenue (RWF,ln) | (11) Savings account | (12) At least one account |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Panel A: IV** | | | | | | | | | | | | |
| Score 5-10 km of mill | -0.012 | 0.023 | 0.010 | 0.099 | -0.062 | 0.226* | 0.010 | 0.009 | -0.050 | -0.522* | 0.022 | 0.013 |
| | (0.013) | (0.024) | (0.023) | (0.099) | (0.060) | (0.134) | (0.010) | (0.015) | (0.055) | (0.271) | (0.022) | (0.018) |
| Score within 5km of mill | 0.001 | -0.003 | 0.011 | 0.036 | 0.143** | 0.116 | 0.017 | 0.015 | 0.045 | 0.455* | -0.010 | -0.003 |
| | (0.014) | (0.026) | (0.022) | (0.099) | (0.069) | (0.218) | (0.011) | (0.017) | (0.053) | (0.267) | (0.022) | (0.017) |
| **Panel B: Competition** | | | | | | | | | | | | |
| Competition | -0.002 | -0.004 | 0.001 | 0.009 | -0.016* | 0.003 | 0.004*** | -0.003 | 0.006 | -0.015 | -0.001 | -0.002 |
| | (0.002) | (0.003) | (0.003) | (0.014) | (0.009) | (0.025) | (0.001) | (0.002) | (0.007) | (0.045) | (0.003) | (0.003) |
| Engineering controls | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Geographic controls | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Mill controls | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Farmer controls | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Adjusted $R^2$ | 0.17 | 0.09 | 0.04 | 0.04 | 0.04 | 0.11 | 0.19 | 0.02 | 0.10 | 0.15 | 0.06 | 0.06 |
| Observations | 1187 | 1187 | 1187 | 1187 | 209 | 207 | 1177 | 1187 | 138 | 137 | 1187 | 1187 |
| **Panel C: IV, No controls** | | | | | | | | | | | | |
| Score 5-10 km of mill | -0.032*** | -0.000 | 0.007 | 0.060 | 0.002 | 0.166* | 0.050*** | -0.010 | 0.012 | -0.425*** | 0.007 | 0.009 |
| | (0.008) | (0.015) | (0.012) | (0.060) | (0.044) | (0.090) | (0.007) | (0.009) | (0.026) | (0.139) | (0.016) | (0.012) |
| Engineering controls | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO |
| Geographic controls | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO |
| Mill controls | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO |
| Farmer controls | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO | NO |
| Adjusted $R^2$ | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.04 | 0.00 | -0.01 | 0.06 | 0.00 | 0.00 |
| Observations | 1210 | 1210 | 1210 | 1211 | 212 | 210 | 1200 | 1210 | 142 | 141 | 1210 | 1210 |
| Dep. var. mean | 0.50 | 0.24 | 0.18 | 0.69 | 7.15 | 4.22 | 6.76 | 0.12 | 5.16 | 10.82 | 0.78 | 0.83 |

*Note:* Bootstrapped standard errors in parentheses. $***$ ($**$) [$*$] indicates significance at the 0.01 (0.05) [0.1] level. These panels provide several raw correlations. Panel A: the instrument (score within 5-10 km) and farmer outcomes from the 2019 farmer survey whilst controlling for suitability in the catchment area (score within 5km of mill). Panel B: between the endogenous variable (competition, number of mills within a 10 km radius) and farmer outcomes from the 2019 farmer survey and Panel C between the instrument (score within 5-10 km) and farmer outcomes from the 2019 farmer survey with no controls. The dependent variables are as follows across the table: column 1 is income from non-coffee source (%); column 2 is the likelihood if the farmer has other sources of income aside coffee; column 3 is the likelihood if they worked for someone; column 4 is the number of days they worked for someone, and in column 5, the wage rate received. Total earnings from outside work is column 6. Column 7 is the (ln) payment they provide for one day of work if they hired someone to work on their farm. Column 8 is the likelihood if they sell milk, column 9 is the price they receive if they sell milk, and column 10 is the total revenue from milk sold. Column 11 is the likelihood if they have a saving account (with a Bank or savings group) and column 12 is the likelihood if they have any type of financial account (bank, savings group, mobile money). Engineering and geographical controls include within a 5 km radius: flexible control of FAO-GAEZ coffee suitability, elevation, slope, spring presence, road density, coffee tree density and rivers. Farmer controls are from the 2019 farmer survey and include distance to nearest mill, age, education, gender, cooperative membership, coffee tree holdings (ln) and latitude and longitudinal coordinates of the farmers' village centroid. Mill controls include cooperative status, mill age, mill age squared and coordinates of the mill. Adjusted $R^2$ is provided for Panel A (IV).

## Table B6: Farmer Outcomes and Market Access

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Dependent Variable | Relational Contract Index | Relational Contract Index | Relational Contract Index | Relational Contract Index | Relational Contract Index | Relational Contract Index | Cherry price (ln) | Share sold as cherries (%) | Home process for saving, dummy | Self financed inputs | Yield (ln) | Job satisfaction index |
| **Panel A: IV** | | | | | | | | | | | | |
| Competition | -0.237*** | -0.213*** | -0.213*** | -0.205*** | -0.236*** | -0.214*** | 0.011*** | -0.104*** | 0.055** | 0.018 | 0.031 | -0.097*** |
| | (0.043) | (0.052) | (0.049) | (0.063) | (0.060) | (0.062) | (0.004) | (0.026) | (0.027) | (0.023) | (0.041) | (0.029) |
| Market Access | | | 0.010 | -0.088 | -0.254** | 0.025 | 0.028** | -0.143*** | 0.176*** | 0.012 | -0.045 | -0.135** |
| | | | (0.065) | (0.057) | (0.103) | (0.062) | (0.011) | (0.054) | (0.052) | (0.064) | (0.101) | (0.062) |
| **Panel B: OLS** | | | | | | | | | | | | |
| Competition | -0.086*** | -0.090*** | -0.089*** | -0.085*** | -0.093*** | -0.090*** | 0.005*** | 0.004 | 0.010 | 0.018** | 0.047*** | -0.036*** |
| | (0.014) | (0.019) | (0.019) | (0.016) | (0.019) | (0.017) | (0.001) | (0.008) | (0.008) | (0.008) | (0.015) | (0.010) |
| Market Access | | | 0.054 | -0.143*** | -0.088 | 0.036 | 0.021 | -0.015 | 0.123*** | 0.010 | -0.026 | -0.064 |
| | | | (0.069) | (0.051) | (0.084) | (0.072) | (0.013) | (0.045) | (0.047) | (0.056) | (0.097) | (0.047) |
| Sample | Baseline, Full | Reduced | Reduced | Reduced | Reduced | Reduced | Reduced | Reduced | Reduced | Reduced | Reduced | Reduced |
| Market access measure | | | Urban Towns | Capital | Sector Capital | Trade Center | Sector Capital | Sector Capital | Sector Capital | Sector Capital | Sector Capital | Sector Capital |
| Score within 5 km of mill | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Geographic controls | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Mill controls | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Farmer controls | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Adjusted $R^2$ | 0.07 | 0.11 | 0.11 | 0.12 | 0.09 | 0.11 | 0.24 | . | 0.04 | 0.07 | 0.19 | 0.07 |
| Observations | 869 | 606 | 606 | 606 | 606 | 606 | 606 | 600 | 606 | 606 | 603 | 596 |

*Note:* Bootstrapped standard errors in parentheses. $***$ ($**$) [$*$] indicates significance at the 0.01 (0.05) [0.1] level. Column 1 is our baseline specification for farmer-specific relational contract index (Table III, column 11). Column 2 is the baseline specification (as in column 1), with the sample of farmers for whom we have village location and as a result the sample is labelled as "Reduced". Note the farmer relational contract index is not subsequently "re-standardized" due to the reduced sample. Columns 3 to 6 are the same specifications as column 2 but now include an additional control variable: a measure of farmer-specific market access. We measure market access following Donaldson and Hornbeck (2016), $\sum p_d/d_{od}^{\epsilon}$ where $p_d$ is the population of the destination point and $d_{od}$ is the distance between origin and destination. Origin in all specifications are the farmer's village centroid coordinates and the destination coordinates varies across the columns depending on the chosen market destination. The elasticity measuring how trade volumes fall as travel times increases, $\epsilon$, is set at 2 for all market access measures (results are robust to different values of $\epsilon$). All market access measures use population data from the *2012 Rwanda Population and Housing Census* available at `http://statistics.gov.rw/datasource/42`, accessed November 2019. In column 3 the market access measure is $\sum p_u/d_u^2$ where $p_u$ is urban town populations (provided by the 2012 Rwanda Population and Housing Census) and $d_u$ is the road distance between the origin and urban towns. In column 4 the market access measure is $\sum p_k/d_k^2$ where $p_k$ is the population of the capital city (Kigali, population provided by the national census) and $d_k$ is the road distance between the origin and capital city. In column 5 the market access measure is $\sum p_s/d_s^2$ where $p_s$ is the population of the sector capitals (provided by the national census) and $d_s$ is the road distance between the origin and the official sector capital. In column 6 the market access measure is $\sum p_m/d_{t_m}^2$ where $p_m$ is the population of the market centers and $d_{t_m}$ is the road distance between the origin and the market trading centers. Note, market centers' do not have a population measure in the national census hence we split evenly the sector's population to all the market centers within the sector. For farmer outcomes in columns 7 to 12, market access destination is defined as the sector capital. Geographic, mill and farmer controls are as defined previously in the baseline. Adjusted $R^2$ is provided for Panel A (IV).

## Table B7: Robustness of Main Mill and Farmer Results

| Dependent Variable | (1) Total Unit Cost (RWF,ln) | (2) Total Processing Unit Cost (RWF,ln) | (3) Unit Labour Costs (RWF,ln) | (4) Days with too many workers | (5) Days with too few workers | (6) Overall Quality Score | (7) Cherry price (RWF ,ln) | (8) Home process for saving, dummy | (9) Self financed inputs | (10) Job satisfaction index | (11) Share sold as cherries (%) | (12) Farmer revenues (RWF,ln) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Panel A: Baseline** | | | | | | | | | | | | |
| Competition | 0.046*** | 0.071*** | 0.107** | 0.227*** | 0.095* | -0.146* | 0.012*** | 0.059*** | 0.034* | -0.059** | -0.075*** | -0.081* |
| | (0.015) | (0.027) | (0.052) | (0.082) | (0.053) | (0.078) | (0.003) | (0.020) | (0.018) | (0.024) | (0.018) | (0.049) |
| Adjusted $R^2$ | -0.11 | 0.03 | 0.24 | 0.22 | 0.11 | 0.04 | 0.26 | 0.00 | 0.08 | 0.12 | . | 0.36 |
| **Panel B: No roads in IV** | | | | | | | | | | | | |
| Competition | 0.048*** | 0.072*** | 0.104* | 0.232*** | 0.086 | -0.144* | 0.013*** | 0.060*** | 0.035* | -0.055** | -0.083*** | -0.095* |
| | (0.016) | (0.028) | (0.054) | (0.087) | (0.054) | (0.080) | (0.003) | (0.020) | (0.021) | (0.026) | (0.015) | (0.056) |
| Adjusted $R^2$ | -0.13 | 0.02 | 0.24 | 0.21 | 0.12 | 0.04 | 0.26 | -0.00 | 0.08 | 0.12 | . | 0.35 |
| **Panel C: Trees 1999** | | | | | | | | | | | | |
| Competition | 0.044** | 0.095** | 0.159** | 0.288** | 0.152* | -0.189 | -0.003 | 0.081** | 0.094*** | -0.068** | -0.120*** | -0.195*** |
| | (0.021) | (0.048) | (0.077) | (0.130) | (0.081) | (0.167) | (0.005) | (0.032) | (0.035) | (0.031) | (0.027) | (0.072) |
| Adjusted $R^2$ | -0.07 | -0.12 | 0.14 | 0.10 | 0.02 | -0.05 | 0.26 | . | . | 0.10 | . | 0.23 |
| **Panel D: Alternative Competition Measure** | | | | | | | | | | | | |
| Competition | 0.199*** | 0.302** | 0.463* | 0.995** | 0.417* | -0.659* | 0.047*** | 0.233*** | 0.133 | -0.229*** | -0.292*** | -0.314* |
| | (0.074) | (0.125) | (0.251) | (0.391) | (0.252) | (0.379) | (0.011) | (0.081) | (0.081) | (0.086) | (0.069) | (0.176) |
| Adjusted $R^2$ | -1.18 | -0.61 | -0.24 | -0.44 | -0.17 | -0.11 | 0.19 | . | -0.02 | -0.00 | . | 0.33 |
| Score within 5 km of mill | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Engineering controls | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Geographic controls | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Mill controls | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES | YES |
| Farmer controls | – | – | – | – | – | – | YES | YES | YES | YES | YES | YES |
| Observations | 177 | 176 | 176 | 173 | 174 | 158 | 869 | 869 | 869 | 856 | 860 | 848 |
| Dep.var. mean | 7.48 | 6.51 | 5.44 | 2.20 | 1.29 | - | 5.33 | 0.23 | 0.35 | - | 0.79 | 4.84 |

*Note:* Panel A are the baseline model results reported for the mill (Table IV - columns 1, 5, 10, 8 and 9, Table V - column 1) and farmers (Table VI - columns 1, 4, 6, 10, 3 and 9). Panel B is the specification in Table B3 column 2; in this specification, the presence of roads is removed in the engineering model when constructing the instrument. That is, there are no roads in the construction of the suitability score in the donut for this specification. Panel C is the specification in Table B3 column 8. In this specification, the engineering model estimation uses the FAO-GAEZ coffee suitability instead of the 30,000 coffee trees criterion. Further, coffee trees are extracted from the 1999 National Coffee Census in the instrument, catchment suitability score and geographic controls. Panel D is the specification in Table B8 column 7. Here, we normalize the number of competing mills by the number of coffee trees (in '000) and also normalize the number of competing mills in the 10 km radius by the total number of coffee trees ('000s, from the 2009 National Coffee Census) in the mills' catchment radius of 5 km. Standard errors adjusted for arbitrary spatial clustering using the *acreg* package written by Konig et al. (2017) are reported for the main mill-level outcomes results in columns 1 to 6. Bootstrapped standard errors in parentheses are reported for the main farmer-level outcomes in columns 7 to 12. $***$ ($**$) [$*$] indicates significance at the 0.01 (0.05) [0.1] level. All specifications are the 2SLS model. Adjusted $R^2$ is provided for all the panels.

Table B8: Robustness to Definitions - Competition and Catchment Area

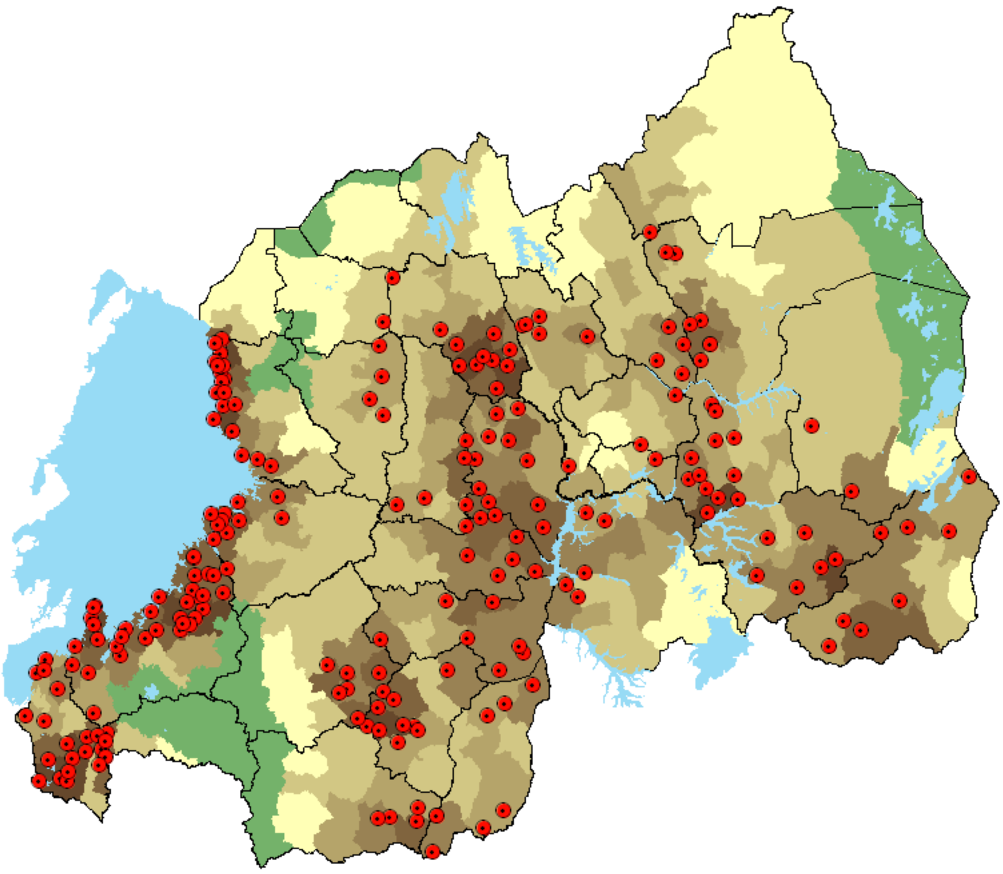
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
|---|---|---|---|---|---|---|---|---|---|---|
| Dependent Variable | Relational Contract Index | | | | | | | | | |
| | Baseline | Catchment size 3km | Catchment size 4km | Catchment size 6km | Catchment size 7km | Mill size 3-5-7km | Alternative competition measure | | | |
| **Panel A: IV** | | | | | | | | | | |
| Competition | -0.283*** | -0.543*** | -0.640*** | -0.195*** | -0.098** | -0.370** | -1.219*** | -1.014** | -0.923** | -1.852** |
| | (0.072) | (0.151) | (0.232) | (0.061) | (0.043) | (0.163) | (0.385) | (0.421) | (0.441) | (0.913) |
| **Panel B: OLS** | | | | | | | | | | |
| Score within 5-10 km of mill | 1.610*** | 1.041*** | 0.826*** | 1.843*** | 2.739*** | 1.308*** | 0.373*** | 0.274*** | 0.255** | 0.247** |
| | (0.312) | (0.225) | (0.288) | (0.347) | (0.431) | (0.492) | (0.089) | (0.097) | (0.102) | (0.110) |
| Score within 5 km of mill | 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 |
| Centered $R^2$ | 0.19 | 0.01 | -0.50 | 0.29 | 0.34 | -0.54 | -0.71 | -0.38 | -0.26 | -2.32 |
| Observations | 175 | 175 | 175 | 175 | 175 | 175 | 175 | 175 | 175 | 173 |

*Note:* Standard errors adjusted for arbitrary spatial clustering using the *acreg* package written by Konig et al. (2017). $***$ ($**$) [$*$] indicates significance at the 0.01 (0.05) [0.1] level. Dependent variable is the relational contract index (z-score). Column 1 is the baseline specification using the engineering model (Table B2, column 4). Columns 2 to 5 change the baseline radii definition of the mill from 5 km (column 1) to 3 km (column 2), 4 km (column 3), 6 km (column 4) and 7 km (column 5). In column 6, a mill-specific catchment size is defined based on the size of the mills' installed capacity (measured in tons of capacity to process cherries). A small size mill's (up to 250 tonnes capacity) catchment radius is defined at 3 km, medium size mill's (between 250 to 500 tonnes processing capacity) catchment radius at 5 km and a large size mill's (more than 500 tonnes capacity) radius at 7 km. In column 7 we normalize the number of competing mills by the number of coffee trees ('000s from the 2009 National Coffee Census). In column 8 we use this alternative measure of competition (column 7) but also remove tree density as a control; note that coffee trees are not used in computing the suitability score of the catchment area. In column 9, the alternative normalized definition of competition (column 7) is used; in contrast, trees are used neither in the catchment area suitability score nor the instrument score (donut area). Column 10 defines competition as the total installed capacity of competing mills (baseline definition) over the potential production of cherries in the catchment area. Potential production is computed as $PP_i = D_i \times \pi \times r^2 \times y$ with $D_i$ tree density (in '000) in the mill's catchment area, $r = 5km$ the (baseline) catchment area radious, and $y = 3.5Kg$ a conservative yield factor of cherries per coffee tree. Centered $R^2$ is provided for Panel A (IV).

# C Additional Figures

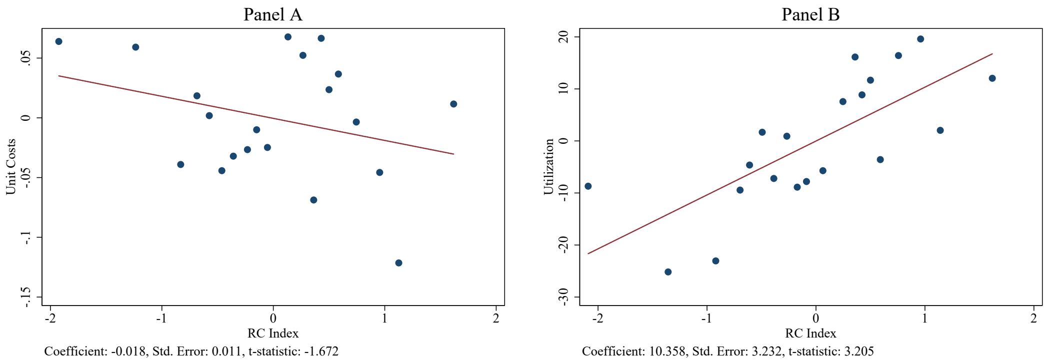
## List of Figures

Figure C1: Mill Placement in Rwanda, 2012



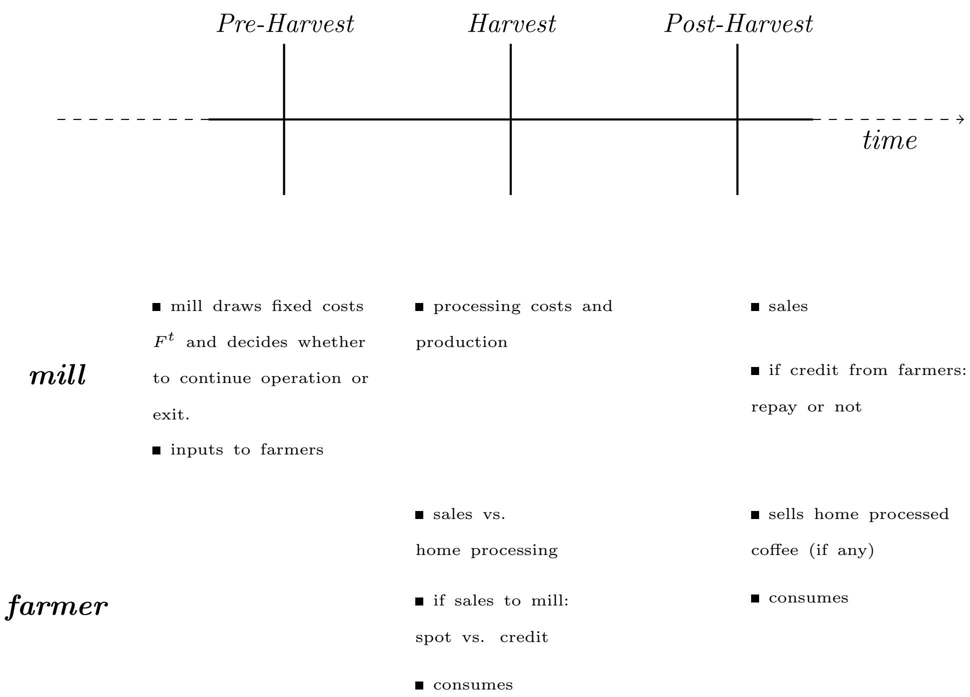
*Note:* This figure illustrates in 2012 the spatial distribution of mills in Rwanda denoted by red dots. In the 2012 harvest season there were in total 214 mills of which 197 were operating. Green shade indicates national parks and conservation areas. Blue shade indicates water bodies. The background overlay in brown is the number of coffee trees at the sector level (the third administrative unit of Rwanda). The darker the shade of brown the higher the number of coffee trees in that sector. This figure is for illustration purposes only.

Figure C2: Validating the Relational Contract Index with Unit Cost of Processing and Utilization of Mill



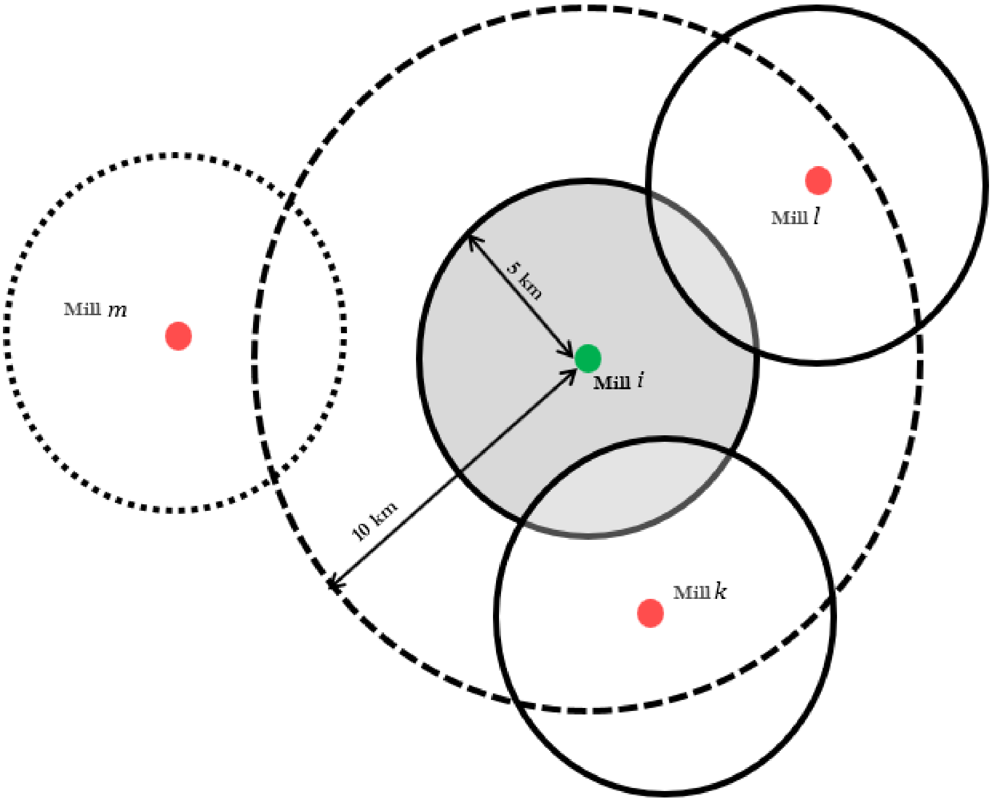

*Note:* Binned scatter plot of mill-level regressions. All regressions control for NGO-support, cooperative status, mill age, mill age squared, and mill coordinates. Controls also include average engineering suitability score, average spring presence, road density, tree density, rivers, flexible control of FAO-GAEZ coffee suitability, elevation, and slope, all within 5 km of the mill. The RC Index is an aggregate of farmer- and mill manager-based indicators of mill-provided inputs, of second payments, and of post-harvest loans. Total unit costs are operating costs (in Rwandan Francs) per kilogram of output produced (parchment). Capacity utilization is cherries processed (tons) in the season divided by the theoretical installed capacity (tons) reported to the coffee board.

Figure C3: Timing of Events During a Coffee Season



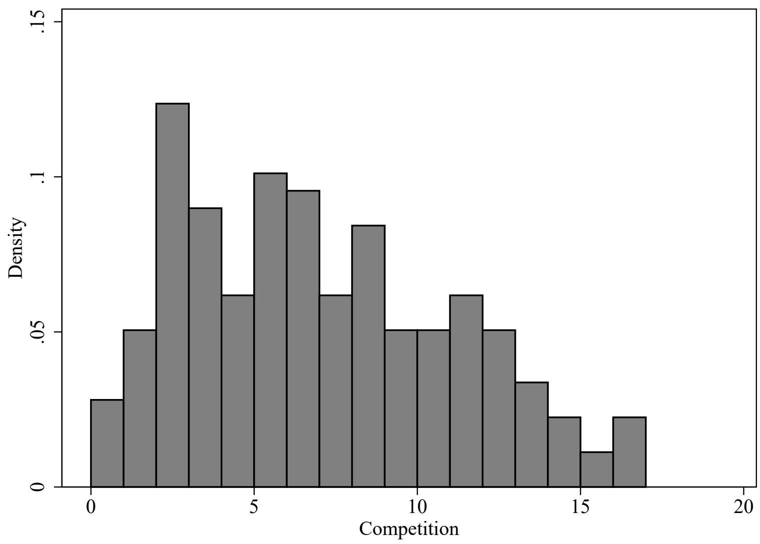

*Note:* This figure depicts the timing of events in the model. Time is an infinite sequence of identical seasons. Each season is divided into three sub-periods: pre-harvest, harvest, and post-harvest. Prior to harvest, the mill decides whether to provide inputs to farmers or not. Farmers then decide 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 the farmer or default. The farmer, after receiving payments from the mills and/or sales from stored, home processed coffee, consumers her income.

Figure C4: Graphical Representation of Competition Measure



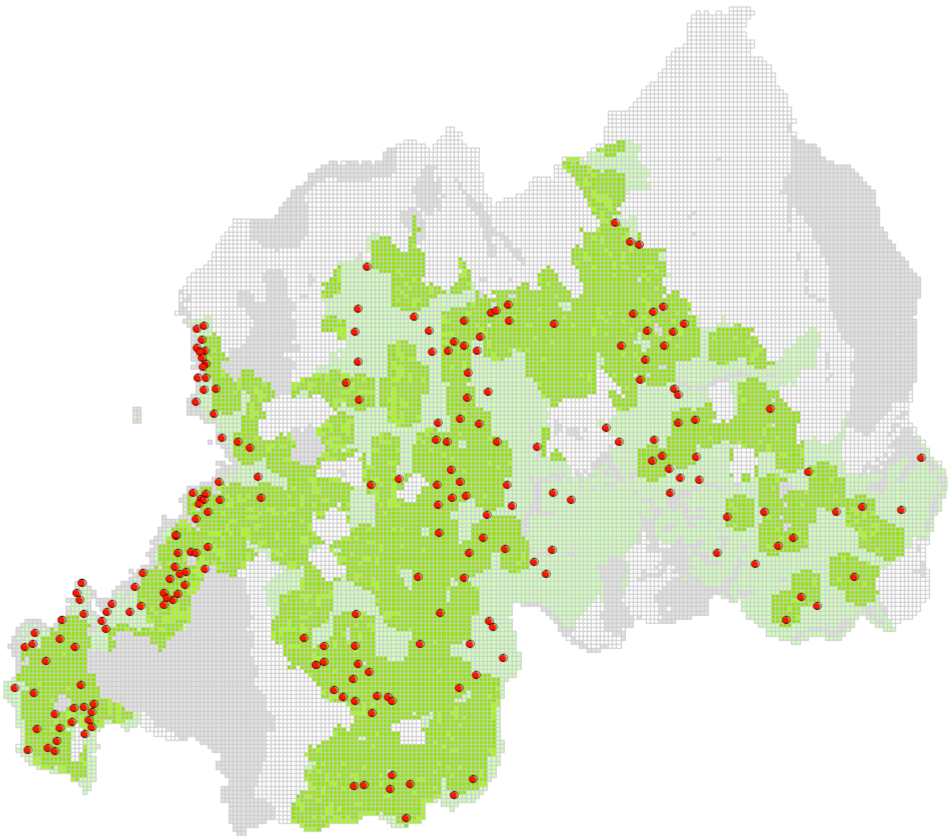

*Note:* This figure illustrates the 5 km catchment area for mill $i$. Any mill within a 10 km radius of mill $i$ will have a catchment area that overlaps (at least to some extent) with mill $i$'s catchment area. The overlap is illustrated in the figure for mill $k$ and $l$. Our competition measure based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

Figure C5: Competition between Mills within 10 km radius



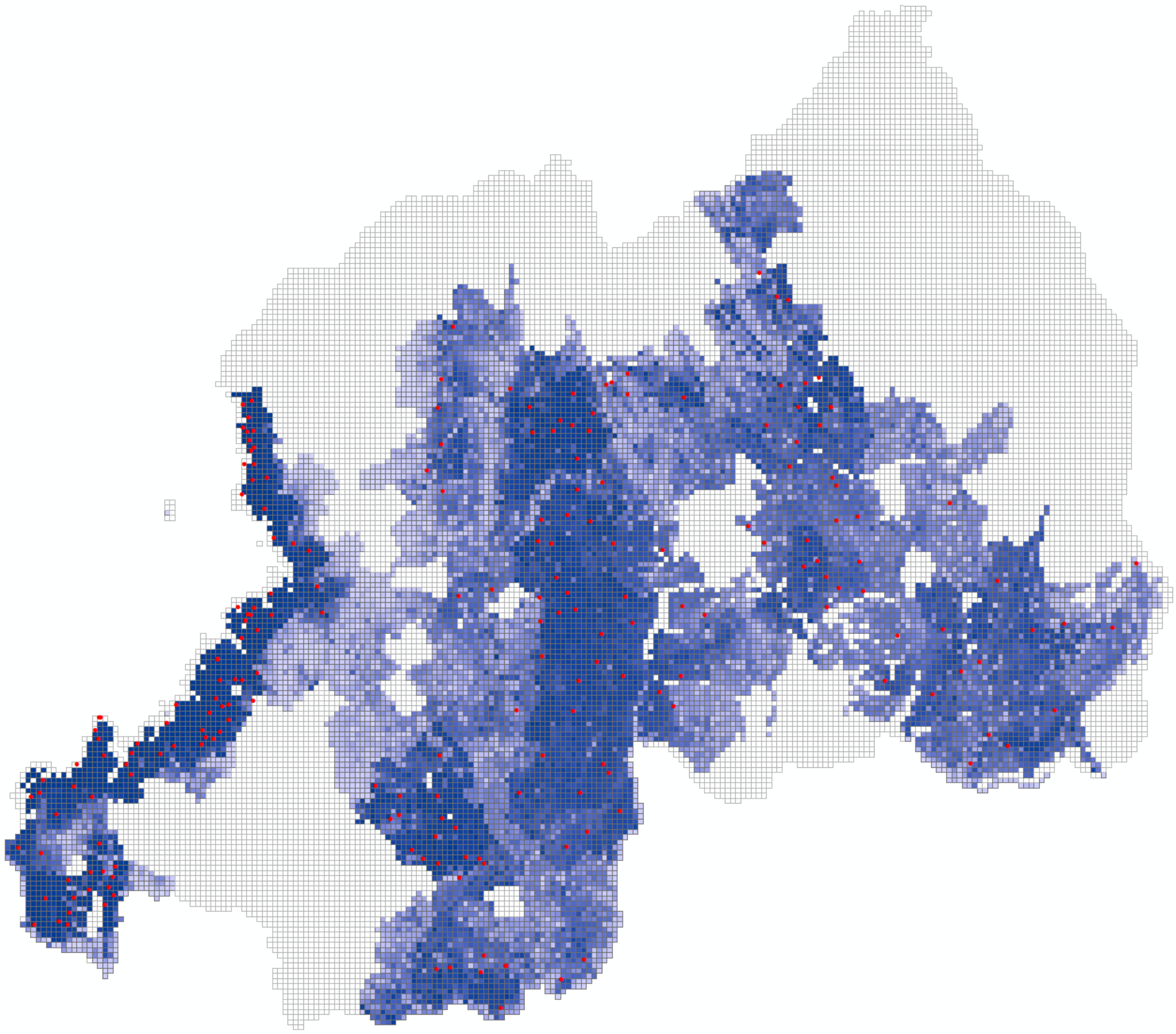

*Note:* The figure plots the distribution of the number of competing mills within a 10 km radius of each mill.

Figure C6: Engineering Model Criteria and Mill Placement



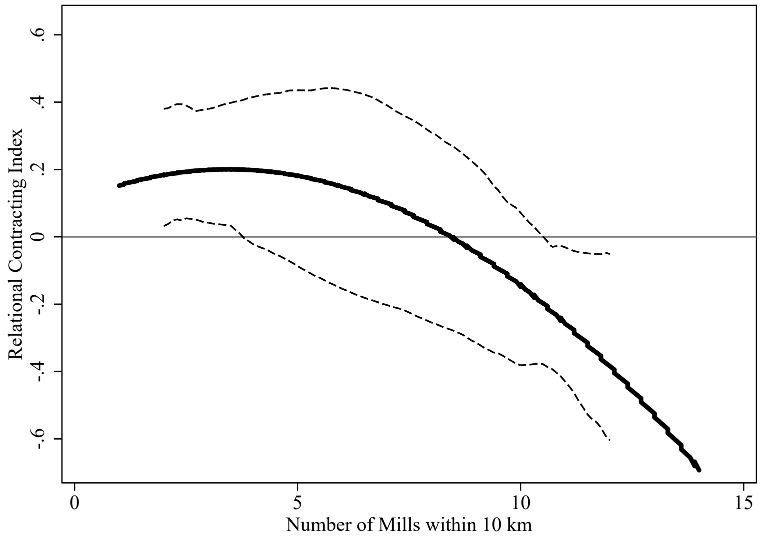
*Note:* This figure illustrates the engineering model's optimal mill placement criteria: the dark grey grids are ineligible for mill placement due to presence of national parks, conservation areas, water body or are built-up areas. The lightest green grids satisfy the number of trees necessary for mill placement while the brightest green grids indicate where the grids satisfy all the criteria (trees, availability of spring water and presence of roads) for mill placement. Red dots depict presence of a mill.

Figure C7: Suitability Score for Mill Placement



*Note:* This figure illustrates the predicted "suitability score" for the placement of a mill in each grid box (1 $km^2$) in Rwanda using our model of mill placement, which is driven by engineering and geographical considerations for the optimal placement of mills. Darker colored grids depict higher suitability score for mill placement. Empty white grids are non-suitable grids for mill placement. Red dots indicate an actual mill location in 2012.

Figure C8: Non-parametric IV




*Note:* 95% confidence interval, represented by dotted-lines, is based on bootstrap re-sampling. Non-parametric IV follows Hall and Horowitz (2005) as implemented by *npivreg* authored by Chetverikov et al. (2018). 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, mill age, mill age squared, and mill coordinates. Controls also include average engineering suitability score, average spring presence, road density, tree density, rivers, flexible control of FAO-GAEZ coffee suitability, elevation, and slope, all within 5 km of the mill. The Relational Contracting index is an aggregate of farmer- and mill manager-based indicators of mill-provided inputs, of second payments and of post-harvest loans. Farmer and mill manager responses are equally weighted.

# D Additional Robustness of Empirical Results

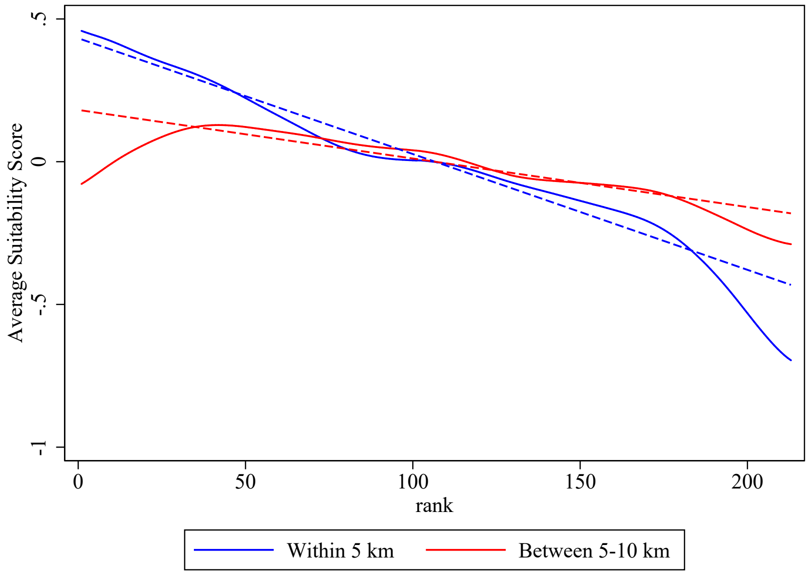
## D.1 Exclusion Restriction

This section discusses additional concerns with regards to the exclusion restriction. Our instrument relies on cross-sectional variation in suitability for mill entry. It is therefore important to consider whether dynamic entry considerations might invalidate the identification strategy. The main threat to the identification strategy is posed by the possibility that certain mills strategically locate in areas with high coffee suitability but that are surrounded by areas with low coffee suitability ("oasis") anticipating lower competition in the future. If those mills are also better at establishing relational contracts with farmers then our identification strategy would be invalid.

To assuage these concerns, we check how the mills' order of entry into the industry 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 a higher suitability score within 5 km from the mill. Under the dynamic strategic entry considerations above we expect earlier entrants to locate in areas with worse surrounding of mill placement. 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 D1 lends support to our identification strategy. The figure confirms that earlier entrants locate in catchment areas with higher suitability for mill placement 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 would be 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.

Numerous field visits and conversations with investors and the coffee board 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

Figure D1: Mill Order of Entry and Suitability Score




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

GIS data, was assembled ex-novo for this article and was thus not accessible before this study to prospective investors and coffee board alike.

A final concern is that mills in more competitive areas might just be worse quality mills along other (unobserved) dimensions. For example, if the cost of entry is lower in those areas, less productive mills (e.g. those with higher operating costs) can survive. Two arguments suggest that the opposite might be the case. First, entry models in which the cost of entry is paid before a productivity draw predict the opposite: when more mills enter, only the best can survive. Second, if surrounding competition matters for the ability to sustain relational contracts, a mill contemplating entering at a location will find it less profitable to enter there conditional on that location's suitability measure. As a result, the mills that do enter will likely be more efficient and more able to sustain relational contracting (e.g., have a higher discount factor) than mills that enter in locations without competition.[40]

[40] If that is the case, the estimated effects might be understated because of selection. Note that this channel – the inclusion of the instrument in a mill entry equation – need not undermine the exclusion of the instrument from the relational contracting equation.

## D.2 Inference

We now consider different assumptions on the structure of the error term in equation (7). There are two main concerns: (*i*) the instrument is the average of a grid suitability score, a predicted variable; (*ii*) errors could be spatially correlated across mills. Table II reports confidence interval under three 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 grid points to predict the score at the grid level for the mill-level regression. This procedure yields the most conservative standard error for the main coefficient of interest (0.088, 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 0.075 and 0.068 respectively. To simplify exposition, in our main analysis we only report estimates with the most conservative method.

## D.3 Definition of Catchment Area and Competition

In our main analysis we defined the catchment area as having a radius of 5 km around the mill and competition to a given mill as the number of mills whose catchment area potentially overlaps with its catchment area. Table B8 explores the robustness of our results to alternative definitions of catchment areas and competition.

Column (1) in Table B8 reports, for ease of comparison, our baseline result. Columns (2) through (6) explore alternative definitions of catchment areas. Columns (2) through (5) show that the results are qualitatively robust as we consider catchment areas of different sizes, 3, 4, 6 and 7 km radius: the impact of an additional competitor on the relational contract index is always negative and statistically significant. Note that as we shrink the definition of catchment areas, the IV identifies the impact of an additional competitor that gets closer and closer to the mill. Reassuringly, the magnitude of our coefficient almost monotonically decreases as we expand the radius of the catchment area from 3 km's to 7 km's; 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. Our baseline estimate thus comfortably lies in the middle of this range of estimates.[41] In our main analy-

[41] The baseline definition is preferred as it strikes a good compromise between two conflicting goals. On the one hand, a wider catchment area includes mills that are not actual competitors. On the other

sis we have defined the catchment area to have a constant radius across mills. Mills, however, differ in installed capacity to process coffee cherries and, plausibly, in how far they source coffee cherries from. Column (6) provides a further robustness check in which we define a mill's catchment area to have a radius of either 3, 5 or 7 kms depending on mill's processing capacity. Results are, again, robust to this alternative definition of catchment size.

We have also defined competition to mill $i$ as the number of mills whose catchment area overlaps with that of mill $i$. It would seem natural, however, to define the intensity of competition relative to the availability of coffee cherries in a given locality; areas more suitable for coffee cultivation will have more coffee trees and also attract more competitors. To account for this, our baseline specifications control for the density of trees in a mill's catchment area. However, as noted above, controlling for trees might induce a bad control problem.

Columns (7) through (10) in Table B8 thus explores alternative specifications in which we normalize the number of competing mills by the number of coffee trees (in thousands) in the mill's catchment area. Results confirm our main findings: an increase in competition reduces the relational contracting index between the mill and surrounding farmers. Column (7) simply substitutes our baseline measure of competition with this modified one which accounts for the availability of coffee trees. Column (8) removes tree density as a control and column (9) further removes trees in the engineering entry model used to compute the instrument. Across the three specifications results are similar and still highly statistically significant (p-value $< 0.05$). Note that although the coefficient is not directly comparable to the baseline, the implied magnitude of the effect is. Column (10) defines competition as the total installed capacity of competing mills divided by tree density in the catchment area and finds similar results.

### D.4 Business Stealing and Scale Effects

The negative impact of competition on mill's outcomes could reflect scale effects induced by business stealing, as in Mankiw and Whinston (1986). Mankiw and Whinston (1986) model an industry in which firms incur fixed costs to enter and then compete over consumers. An additional entrant reduces incumbents' production volume (business stealing) and induces excessive entry through the inefficient duplication of fixed costs. The framework is relevant for, and can be applied to, our context in which coffee mills incur fixed costs and compete over farmers. It is thus important to understand

hand, a narrower catchment area is more likely to violate the exclusion restriction.

how our results differ from this mechanism.

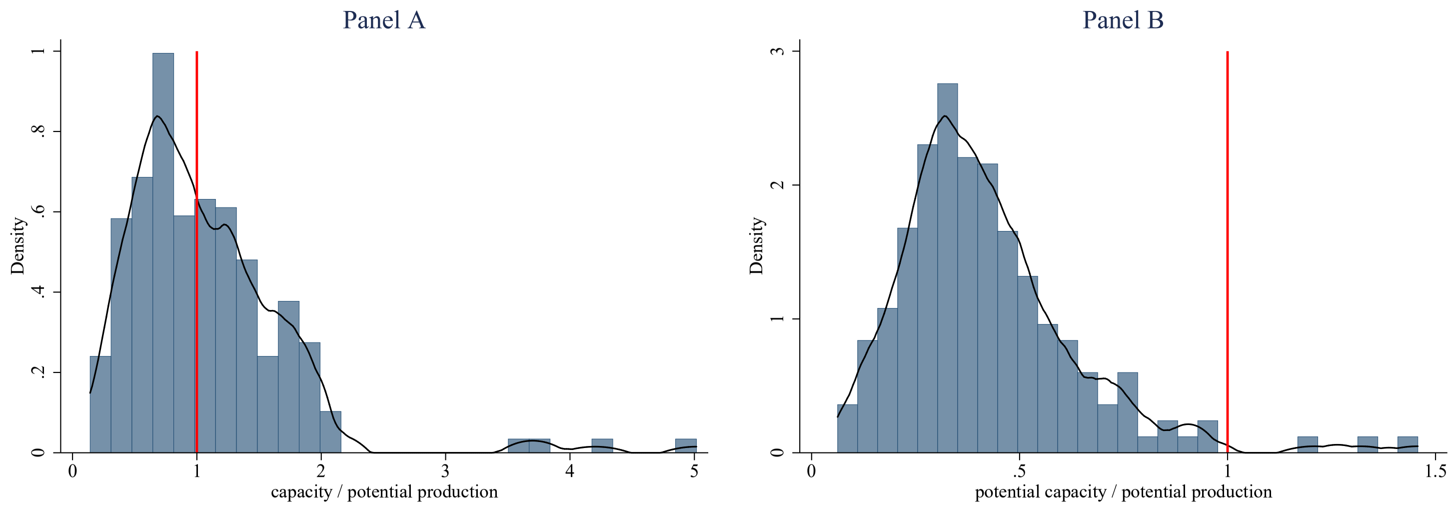
When applied to our context, the model in Mankiw and Whinston (1986) implies that the (inefficient) entry of the marginal mill ($i$) increases the aggregate volume of coffee supplied by farmers to mills in the market and ($ii$) makes farmers weakly better off. The evidence in Section 5.3 already suggests that competition reduces farmers' sales to any mills and likely makes farmers worse off.

The predictions of the mechanism in Mankiw and Whinston (1986) are inconsistent with two additional pieces of evidence. First, in order for business stealing to be relevant, mills must compete over scarce coffee supply. In contrast, there is plenty of coffee to be processed. In 2012, the year of our survey, the share of coffee processed by mills was around 33%. This aggregate figure however is not conclusive: it could be that mills are concentrated in regions in which there is scarcity of coffee cherries relative to installed capacity while most of the home processed coffee originates in regions in which mills have not entered for unrelated reasons.

Figure D2, however, suggests that in many local markets in which mills compete for cherries, there is plenty of coffee to accommodate the mills' installed capacity. The figure reports the distribution of the ratio of aggregate installed capacity and potential production across local markets. The unit of observation is a mill and potential production is defined as the number of trees in its catchment area multiplied by a (conservative) yield factor 3.5 kilograms of cherries per tree.

Panel A defines capacity as the sum of capacity across all mills competing with a mill $i$ according to our baseline definition of competition. This provides an implausibly conservative bound because it assumes that all mills competing with mill $i$ source all their cherries in mill's $i$ catchment area. Even then, we find that in roughly just over 50% of catchment areas the ratio is below one, that is, there is more coffee available than installed capacity. When, more plausibly and consistently with our empirical strategy, we weigh the capacity of each competitor by the area of overlap between the two catchment areas, the estimated ratio is below one in 98% of catchment areas (Panel B). That is, almost everywhere there is excess supply of coffee to be processed and thus scarcity of coffee induced by competition should not be a constraint to mills operation.

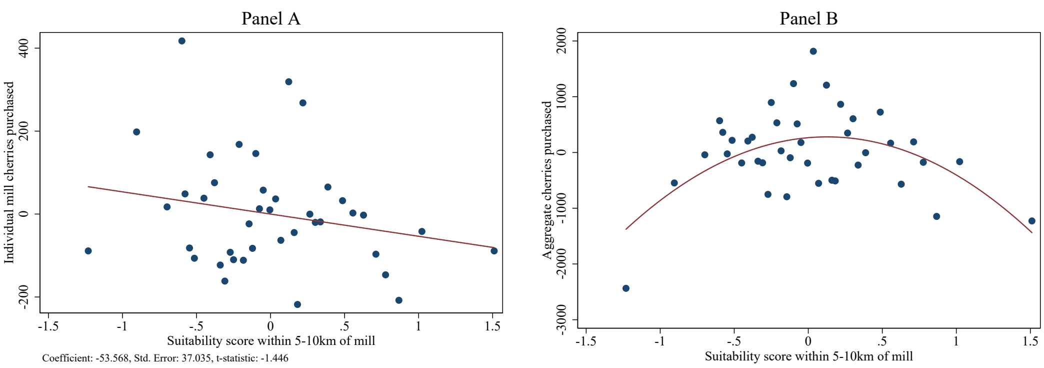
Figure D2: Mill Capacity and Potential Coffee Production




*Note:* In panel A the histogram illustrates the distribution of $\frac{capacity_i}{potentialproduction_i}$ across mills. The unit of observation is a mill $i$, capacity is defined as $c_i + \sum_j c_j$ and $i \neq j$. Other mills in the 10 km radius to mill $i$ are denoted by $j$. Hence, capacity is defined as the sum of capacity across all mills competing with a mill $i$ according to our baseline definition of competition. Potential production is computed as $PP_i = D_i \times \pi \times r^2 \times y$ with $D_i$ tree density (in '000) in the mill's catchment area, $r = 5km$ the (baseline) catchment area radius, and $y = 3.5kg$ a conservative yield factor of cherries per coffee tree. For 53% of the mills, the ratio is below one; that is, there is more coffee available than installed capacity. In panel B, we weigh the capacity of each competitor by the area of overlap between the two catchment areas, or the distribution of $\frac{potentialcapacity_i}{potentialproduction_i}$ for $mill_i$. Potential capacity is defined as $c_i + \sum_j \frac{A_{ij}}{\pi \times 5^2} c_j$, where $A_{ij}$ denotes area of overlap between mills $i$ and $j$. The estimated ratio is below one in 98% of catchment areas, which is to say that almost everywhere there is excess supply of coffee cherries to be processed.

An additional piece of evidence comes from the relationship between competition and aggregate volumes processed by mills. The left panel in Online Appendix Figure D3 shows that our instrument is negatively correlated with the quantity processed by the mill. The figure illustrates the reduced form relationship corresponding to Column 6 in Table IV. *What about the aggregate quantity processed by mills?* The right panel of Figure D3 shows that our instrument displays an inverted-U shaped relationship with the aggregate amount of cherries processed by mill's $i$ competitors. As in our model, past a certain level of entry, aggregate volumes processed decline with further entry.[42]

Figure D3: Cherries Purchased and Suitability Score




*Note:* Panel A (left) is a binned scatter plot of cherries purchased by an individual mill against suitability score (5 to 10 km) with a linear fit. Panel B (right) shows the aggregate quantity which sums up the cherries purchased by competing mills including own. Mill $i$ has competing mill $j$ if mill $j$ is located within 10 km from mill $i$. Panel B shows binned scatter plot of this aggregate quantity against the suitability score with a quadratic fit (semi-parametric). All regressions control for NGO-support, cooperative status, mill age, mill age squared and mill coordinates. Controls also include average engineering suitability score, average spring presence, road density, tree density, rivers, flexible control of FAO-GAEZ coffee suitability, elevation and slope, all within 5 km of the mill.

In sum, we find several pieces of evidence suggesting that competition lowers the aggregate amount of coffee processed by mills. This is consistent with the implications of a model in which competition breaks down relational contracts between mills and farmers and not consistent with outcomes being entirely driven by a business stealing effect in the presence of fixed costs.

[42] Conditional on our baseline set of controls, the correlation between the tons of coffee processed by mill $i$ and the aggregate tons of coffee processed by mill $i$ competitors yields a positive 0.03 coefficient (p-value<0.05). This reflects higher competition in places with more coffee cherries. The non-monotonic relationship in the right panel prevents us from running an IV specification.

## Online Appendix


Chetverikov, D., D. Kim, and D. Wilhelm (2018) “Nonparametric instrumental variable estimation,” *Stata Journal*, Vol. 18, No. 4, pp. 937–950.

Conley, T. (1999) “GMM Estimation with Cross Sectional Dependence,” *Journal of Econometrics*, Vol. 92, No. 1, pp. 1–45.

Donaldson, D. and R. Hornbeck (2016) “Railroads and American Economic Growth: A ‘Market-Access’ Approach,” *Quarterly Journal of Economics*, Vol. 131, No. 2, pp. 799–858.

Konig, M. D., D. Rohner, M. Thoenig, and F. Zilibotti (2017) “Networks in Conflict: Theory and Evidence From the Great War of Africa,” *Econometrica*, Vol. 85, No. 4, pp. 1093–1132.

Mankiw, N. and M. D. Whinston (1986) “Free Entry and Social Inefficiency,” *The RAND Journal of Economics*, Vol. 17, No. 1, pp. 48–58.