Market failures, task-based production model and the agricultural productivity gap

Nicolás Garcés Rodríguez *

June 5, 2023

Abstract

The agricultural productivity gap is one of the main issues of developing economies. The methodologies used to assess these productivity levels are heavily reliant on assumptions over the production function, that, most of the time, do not take into account the heterogeneity of production and factor roles throughout the production process. In this paper, we will take a new approach to agricultural production through the lens of a task-based production model. The empirical strategy takes a novel perspective on market failure analysis by separating market dynamics at each production stage and having a holistic approach to all the production factors, taking the case of coffee farms in rural Colombia. The paper shows that labor and machine demands are biased by household characteristics, proving the existence of market failures in the economy and that such behaviors are more common in farms led by females due to credit market dynamics. Results also hint at the dependency of equilibrium factor allocation on other factors' markets and scales, which potentially causes a double productivity gap when facing a single market failure.

1 Introduction

Agricultural productivity disparities between rich and poor countries play a central role in economic development studies(Restuccia et al. 2008; Gollin et al. 2014). The misallocation of production factors and market inefficiencies are commonly argued to be one of the main causes of the productivity gap, as both generate the accumulation of factors in production settings with low performance and skew agriculture yield downwards (Restuccia and Santaeulalia-Llopis 2017;Adamopoulos et al. 2022,Ayerst et al. 2020;Gollin and Udry 2021).

There is still a lot of discussion about the size and the causes of the productivity gap. The complexity of estimating productivity, due to the multiple methodological approaches that can be taken, and the assumptions over production functional forms can lead to sharply different conclusions (Bils et al. 2021;Gollin et al. 2014; Chen et al. 2017;Hamory et al. 2021; Gollin and Udry 2021). One common practice, however, is focusing on the scale of production factors. The usual assumptions of Hicks's neutral productivity and substitutability between inputs lead to allocation analysis based on the performance in aggregated production levels, either between sectors or within production units in agriculture.

Understanding factor performance based on the aggregated levels in production is limiting, especially for agriculture. The different dynamics and tasks that agricultural seasons have, create heterogeneous settings where inputs play different roles across the production process. For factor allocation and productivity analysis, this could imply that not only factor levels are relevant for performance, but also the allocation of factors within the multiple stages of production. For market dynamics, on the other hand, the analysis should take into account the seasons where factors are more intensive and, therefore, use a disaggregated approach in order to account for this heterogeneity.

In this paper, we will take a new approach to agricultural productivity gap inspired by the task-based model proposed by Acemoglu & Restrepo (2018). By understanding agricultural production

*Universidad del Rosario. ORCID:0000-0003-3940-1453; e-mail:nicolas.garces@urosario.edu.co. Special thanks to Paul Rodríguez and Mauricio Rodríguez for the guidance and excellent support as supervisors of this project. Thanks to Alianza EFI for providing me access to the data.

as a series of tasks, the theoretical framework expands the productivity understanding by separating production stages into Harvest and Non-Harvest tasks, assessing the performance of labor and capital subject to each stage of production. The potential causes of the productivity gap under this approach, therefore, highlight the heterogeneity across stages and the substitutability of factors relative to the task and the role of each factor.

Our empirical strategy will follow the intuition of the model and analyze market dynamics at separate production stages and heterogeneous characteristics. A holistic analysis of production factor markets will allow us to capture the dynamics between inputs and within production stages. Following Dillon and Barrett (2017) a separability test at each season, accompanied by multiple estimations for labor allocation, capital investment, and credit take-up, will allow us to capture the existence of market failures and the ways farms endowments affect factor demands. We find that both labor and machine demand are biased by households' endowments, meaning that there are market failures that affect resource allocation, as well as heterogeneity in the composition of farms with inefficient allocations due to credit access differences between male and female-led farms. Lastly, we find that misallocation of labor tends to be accompanied by land scale constraints, meaning that one market inefficiency can derive in double productivity gaps.

We will take advantage of a novel microdata survey of coffee producers in Colombia. This crop is extremely important in the agricultural sector and national current account (Ocampo 2015) and thus, analyzing it can give meaningful insights into the entire Colombian agricultural sector. Furthermore, coffee production is no exception to the agricultural productivity gap, as recent estimations show that Colombian coffee farm production is, on average, half of what is feasible with the available resources (Garcés Rodríguez and Gomez 2022). The settings for coffee production, as farms specialize in the crop, and the granular level of the survey makes auspicious the study of productivity in this setting.

The paper is structured as follows. Section 2 will approach the theoretical model and expose the main components in order to settle equilibrium production output within this production dynamics. Section 3 defines the potential causes of misallocation under this scope. In section 5, the empirical strategy will be proposed in order to analyze the framework within the dataset. Section 6 will have a discussion over the results of the model and, lastly, some concluding remarks will be made in section 7.

2 Model

2.1 Production

Based on the Acemoglu & Restrepo model (2018), a task-based production model will be adopted for the agriculture sector. The model considers an agricultural production function based on two tasks: Harvest and Non-Harvest task. Farms produce the final outcome as a combination of these two activities. Each task uses capital, labor, and land, the last one being a fixed factor for each farm. Lastly, tasks are produced separately and sequentially across two production stages. The production function takes the following form

$$Y_i = \sum_{j=1}^{2} y_i(j)^{\sigma_j} \tag{1}$$

Where Y_i is the total gross output of farm i, y_i(j) are tasks productions, where j $\in$ { Harvest, Non-Harvest}, and $\sigma_j \in [0,1]$ is the elasticity of substitution between tasks, with $\sigma_H + \sigma_{NH} = 1$. Capital and labor are substitutes and is assumed that each factor has a relative advantage in one of the tasks. Task production is defined by the following equation

$$y_i(j) = T_i^{\eta_j} \left(\beta(j)K_i(j) + \alpha(j)L_i(j)\right)^{(1-\eta_j)} \tag{2}$$

Where $\eta_j \in$ (0,1) is the share parameter of factors in each task, T_i is a fixed level of the land, $L_i(j)$ and $K_i(j)$ are land and capital levels at each stage, $\beta(j)$ and $\alpha(j)$ are capital and labor productivity respectively in task j. For a given, and exogenous, level of T_i ($\bar{T}_i$) in each farm, the optimal level of the combination of capital and labor in each task is given by the FOC.

$$\bar{T}_i \frac{(1-\eta_j)}{\eta_j} = \beta(j) K_i + \alpha(j) L_i \tag{3}$$

This means that tasks can be taken to optimal levels with a combination of land with both labor and/or machines, depending on their relative productivity on each task. Although the assumption of fixed and exogenous land is strong in agriculture, it allows us to focus the analysis on the factors with a more dynamic supply and more regular demand. Lastly, let there be a comparative advantage of factors in each task such that

$$\beta(\mathrm{j}) \begin{cases} \beta \text{ for j = Harvest} \\ \beta^* \text{ for j = Non-Harvest} \end{cases} \qquad \alpha^*(\mathrm{j}) \begin{cases} \alpha^* \text{ for j = Harvest} \\ \alpha \text{ for j = Non-Harvest} \end{cases}$$

Where $\beta^* > \alpha$ and $\alpha^* > \beta$.

Hence, it is assumed that **Capital is more productive in non-harvest tasks while labor is more productive in harvest activities**. The feasibility of this assumption and the overall fit of the model to coffee production will be discussed in section 4.

2.2 Households

Households follow a basic agricultural framework, as proposed by Dillon & Barrett (Dillon and Barrett 2017) and Barrett, Sherldund & Adesina (Barrett et al. 2008), that captures the process of factor allocation and the maximization of both utility and profits. The households maximize a strictly increasing, concave utility function (U(L^l, C|Z)) that depends on preferences over leisure (L^l), the consumption of goods (C), and is conditional to household characteristics (Z). At each production stage, farms have a total time endowments denoted as $\bar{L}$, that can be used in leisure (L^l), workforce for the farm production (L^F) or it can be supplied in the labor market (L^M) such that

$$\bar{L}(j) = L^l(j) + L^H(j) + L^M(j)$$

The overall labor levels will be defined by a simple leisure demand rule ($U_L'^{\,l}/U_C' = \omega/P_c$), however, labor allocation will depend on relative productivity between sectors. More specifically, we assume that, while family agricultural labor is a perfect substitute for hired labor (no skill needed for agricultural tasks), non-agricultural wages depend on individual skills (S_i) and the economic environment (v) such that $\omega^M = F(S_i, v)$. Thus, based on their skills, household members would choose between being an unpaid worker on the farm production, boosting overall agricultural income, or working outside the farm in non-agricultural activities, that, under competitive market conditions, would imply choosing the maximum benefits of labor($Max = \{\omega^F, \omega^M\}$)

The nonmarketed aspect of household labor in agricultural production is crucial in the characterization of rural households. Although labor is allocated based on skills, this wage-free nature of self-employment creates a strong relationship between household utility and profit maximization. Ultimately, agricultural profits will always be in consideration when choosing optimal labor allocation among household members. Any adverse condition on agriculture (climate shocks, high wages, low prices, etc) that drives production down can eventually be covered by household members' labor because the opportunity cost of not having agricultural output is not minor for the aggregated household income. Conversely, ideal conditions for agricultural production, especially competitive markets, will allow an efficient labor allocation between sectors, that ultimately maximizes both utility and profits.

Lastly, households will also have an endowment of machinery that can be used in agricultural production (K_i^0). Unlike labor endowments, there is a cost for using machinery (Maintenance, electricity cost, etc) denoted as Ω. Before facing each production stage, households can increase their capital levels if needed by investing in more machinery (I), implying that the effective capital used in production would be as follows.

$$K_i = K_i^0 + I_i \tag{4}$$

2.3 Finance

In order to pay wages and invest in machinery, if needed, farms finance their productive activity with credit. The credit market in agriculture is both diverse and extremely complex, but some basic assumptions will be made in order to capture its essence. Although many lenders come into play in this sector, and this usually implies differential interest rates, credit size, and transaction costs, it is assumed that credit lenders are homogeneous, or likewise, that both formal and informal markets complement each other. In terms of quantities, there is also a pronounced heterogeneity depending on markets (Giné 2011), but, ultimately, the observable characteristics of the household, as a measure of their capacity for repayment (or credit profile), will be assumed as the main determinant of the credit levels. Hence, we can define total credit loan values as $q_i = F(X_i|\bar{Q})$, where X_i is a set of observable variables that lenders use in order to assess the capacity of repayment and $\bar{Q}$ is available credit in the market. The empirical strategy will further test if this assumption holds and the feasibility of this characterization for agriculture.

To simplify the model, it is assumed that credit is the only liquidity source for production, as well as it is ruled out any consumption credit, and that, at each production stage, credit availability and costs are independent of the credit in the previous stage. Although the last assumption is quite strong, as it implies the impossibility of moving liquidity through stages as well as differential debt ceilings for a single household at the two production moments, many credit lines, heavily intervened in this type of market, do operate in a way that both cost and loan sizes are different depending the usage of the money (like investment loans v.s crop liquidity loans, for example). Lastly, it will be assumed that the opportunity cost of debt will always be equivalent to production benefits, and thus the present value of the credit will be equal to the production costs.

Credit supply will follow a simple fixed offer, where the total amount of liquidity available to borrowers is denoted as $\bar{Q}$, and $F(X_i)$ will then correspond to a share over available credit. The diversity on X_i, as well as the endowments of machinery and land, creates a heterogeneous credit demand that will turn overall agricultural efficiency into the efficiency of credit supply, as optimal labor and investment must be accompanied by efficient credit allocation as well.

2.4 Production stages

The model assumes that production is sequential in order to capture the seasonality of agriculture. At both stages, households maximize production tasks and their utility, more specifically leisure, as labor choices are taken. Once finished both tasks, the total revue is earned and households take their consumption choices based on benefits and labor incomes. At each stage, independent credit choices are taken in order to finance the productive activity.

The model reduces production in two stages: A Harvest stage, when products are collected, and a Non-Harvest stage, where raw products are transformed. Although we could think of a third stage previous to the harvesting, when land is sown and prepared, the intuition and the model implications remain the same if reduced to two stages. Furthermore, depending on the crop, Non-Harvest could also include the third stage if, at both stages, the assumption over factor productivity advantage is the same.

2.5 Harvest Stage

At the harvest stage, farms collect the crop from the field. Given the short-term requirements of this process, farms will be completely focused on harvesting as much as possible. Farms will minimize the cost of harvesting the crop, which, assuming that is cheaper to use the most productive factor, will imply hiring workers in order to maximize the task. Farms will take loans on the credit markets in order to have the liquidity that allows them to hire the workforce.

Lastly, farms partially maximize their utility as work, and thus leisure, choices take place, both based on the expected revenue of agriculture that has not taken place yet. Following Barrett (2008), the indirect utility maximization function that defines labor allocation at each stage would be as follows.

$$\max_{\bar{L}_i} E[V_i(L^l, P, \omega_H, R, \omega_M | Z_i)]$$
$$\text{S.t}$$
$$\text{Y}_i = \sum_{j=1}^{2} y_i(j)^{\sigma_j}$$
$$\text{L}_i^H = L_H^H + L_H^M$$
$$\bar{L}(H) = L^l(H) + L^H(H) + L^M(H)$$
$$\omega_M = F(S_i, v)$$

2.6 Non-Harvest Stage

At this stage, once the crop has been collected, the farm focus on the transformation processes of the crop. While, usually, there are markets for raw agricultural products, most crops need some level of transformation (like packaging the product) to, at least, be sold on the raw market. The relative productivity advantage of machines over labor would make it the ideal factor to produce this task, however, farms minimize production costs of tasks and, thus, some conditions must hold in order to achieve this outcome.

First, based on their machinery endowments (K_0), farms choose the cheapest factor to produce the task, taking into account observed wages, machines' cost of usage, and relative productivity. In order to have machines as the optimal production factor for the non-harvest tasks, the following condition must hold.

$$R + \Omega \leq \omega_{NH} \frac{\beta^*}{\alpha} \tag{5}$$

Where R is the price of machines, Ω is the cost of using machines, and ω_{NH} is the wages for Non-Harvest labor. Equation 5 is denoted as the ideal price level condition, and simply states the cost efficiency of machinery usage over labor in non-harvest tasks. Beyond marginal advantage, quantities are also important in overall expenses, and, initial endowment levels (K_i^0), might have several implications for efficient choices. More specifically, even if ideal price condition hold, a farm with a low initial machine level might incur greater costs when taking capital up to the optimal levels compared with hiring workers and pairing them with the initial machines. This case could be stated as a condition over K_i^0 by solving the equation cost in each case, where the positive investment condition would require the following.

$$\frac{\bar{T}(1-\eta_{NH})\left(\frac{\Omega+R}{\beta^*} - \frac{\omega_{NH}}{\alpha}\right)}{\eta_{NH}\left(\Omega + R - \omega_{NH}\frac{\beta^*}{\alpha}\right)} \leq K_i^0 \tag{6}$$

if K_i^0 satisfies condition 6, it will be cheaper to invest in machines up to optimal levels than hiring workers. This way, when condition 5 and 6 hold, machines are the optimal choice for non-harvest production in household i, while only satisfying condition 5 would lead to a mix of K_i^0 and labor that satisfies FOC as the optimal choice. Thus, investment levels are defined by a corner solution as follows

$$I_i \begin{cases} (\mathrm{K}^* - K_0) & \text{if condition 6 and 5 hold} \\ 0 & \textit{Otherwise} \end{cases}$$

As this is the final stage of agricultural production, farms will have the revenue of production and will be able to choose the level of consumption. This way, household utility maximization takes the following form.

$$\max_{\bar{L_i},C} U_i(L^l, C|Z_i)$$

$$\text{S.t}$$

$$\text{pC} = \Pi_i + \omega_M L^M$$

$$\bar{L} = L^l + L^H + L^M$$

$$\text{L}^l, L^H, L^M, C \geq 0$$

2.7 Equilibrium

Under these settings, the cost of producing tasks is minimized, which means that, for task j, the price of the factor selected at each task in equilibrium is as follows:

$$P(j) = min\left\{\frac{R+\Omega}{\beta(j)}, \frac{\omega_j}{\alpha(j)}\right\}. \tag{7}$$

If the ideal price condition holds and initial machines endowments are enough for having investment (conditions 5 and 6 holds), the Non-harvest task is fully developed with machines while the Harvest task is done with labor. Thus, farm i demand of labor and capital will be as follows

$$\text{L}_i^* = \tfrac{\bar{T_i}}{\alpha^*}\tfrac{1-\eta_H}{\eta_H} \qquad K_i^* = \tfrac{\bar{T_i}}{\beta^*}\tfrac{1-\eta_{NH}}{\eta_{NH}}$$

These demands must be accompanied by the respective equilibrium credit demand that would take the following form.

$$\text{q}_{NH}^* = (1+r)(R(K_i^* - K_i^0) + \Omega(K_i^*)) \qquad q_H^* = (1+r)(L_i^*\omega_H)$$

3 Productivity gap

Many market or production dynamics can distort farms' choices, deriving in factor allocations that have lower outcomes than what could be feasible in equilibrium. In this section, we will describe and analyze the potential distortions that could arise in the model.

3.1 Inadequate investment conditions

As equation 5 states, the scale of stock machines heavily conditions current investment. A farm that has low levels of capital endowments will not have incentives to buy new machines at any stage of production and, therefore, allocate labor in the task which is not the most efficient factor. The overall distribution of these endowments in the economy could also imply a wide heterogeneity in machine demand and credit needs, especially if such endowments are not in line with land distribution.

The substitutability between labor and machines, the constant need for agricultural output that many rural households have, and the existence of markets for both raw and processed agricultural products create conditions where output can be sustained without investment. Many farms with low, sometimes zero, machinery scale might have no incentives to invest until extraordinary conditions make profitable the investment in machinery. This threshold of production settings without machinery can create a productivity trap in which only external circumstances can change the production nature of farms with initial adverse conditions or low endowments.

Lastly, and sometimes overlooked element, is the machine market in rural sectors. The dispersion of farms and the distance from the urban areas makes it difficult to have good information on the available products for production. Although R&D is responsible for producing affordable machines, the price that producers face when buying them is the ultimate determinant of capital adoption. The relevance of competitive and available supply in rural regions is not minor for the development of investment in machines. It is worth noting that all this discussion makes sense when talking about machines for Non-Harvest tasks. The model highlights the importance of focusing investment and capital accumulation toward a specific class of machines, thus, in order to have equilibrium levels of investment, is not only important to have auspicious conditions, but also directed accumulation of capital.

Proposition 1: *Low capital endowments can lead to long-term unproductive production settings.*

3.2 Labor market constrains

For both stages of production, and even for utility maximization, labor markets play a pivotal role. Workforce supply for agriculture is extremely volatile and heavily pressured during harvest season, as both labor intensity and regional homogeneity would imply a soaring demand at specific moments in time and space. As long as the market is able to satisfy aggregated labor demand, the agricultural sector will have no trouble in producing the tasks that need this factor. However, a limited labor supply is not only an expected situation in modern rural settings (Foster and Rosenzweig 2007) but also a main potential limitation for attaining equilibrium levels.

Let there be a limited labor supply at any stage of production denoted as $\bar{L}$, such that overall labor demand exceeds this level. Beyond any capacity of individual farms for hoarding labor, like having enough resources to offer higher wages or information advantages, it could be assumed that, on average, supply would be deficient, and hired workers would not be enough for equilibrium production. In this situation, the household workforce could be the ideal complement for labor levels if the allocation of this type of labor in agriculture is optimal. To put it formally, the marginal benefit of augmenting labor in task production must be greater than the potential wage that household members would receive in the non-agricultural labor market. Assuming this situation happens in the harvest stage, this condition would be as follows.

$$P_y \left(\frac{\sigma(1-\eta_H)}{L^M_{Hcap}} \right) \geq \omega^M \tag{8}$$

Where P_y is the agricultural commodity price and L^M_{Hcap} is the total hired workers in the market by the farm i, such that $L^M_{cap} < L^*_i$. The capacity for compensating for supply shortage would be then subject to the number of household members as well as their skill levels, where more skills would make less likely the compensation of labor. Ultimately, taking each task, harvest in this case, to optimal levels, would depend on the degree of market constraint (how many workers can be hired) as well as the overall share of the factor in the task. This way, the productivity gap caused by market constraints could, at least partially, be covered.

Up to this point, the labor compensation makes sense. The factor is ideal for the task and having household members working on the harvest task would not imply any inefficiency in terms of the factor allocation. However, the case is the opposite if the labor market failure happens in the Non-Harvest stage. A farm that, inefficiently, allocates labor in this task to pair it with initial machines endowments, might face this labor shortage and thus the condition for labor compensation would then take the following form.

$$P_y \left(\frac{(1-\sigma)(1-\eta_{NH})}{\beta^* K_0 + \alpha L^M_{NHcap}} \right) \geq \omega^M \tag{9}$$

We could think of this case as a double productivity gap. A farm with low machine stock levels that do not pass the threshold for positive investment could then also face labor market constraints in a way that household members would be retained on an agricultural task that should not even be produced with labor. This crossed misallocation, given by the substitutability of factors, could be the cause of the unproductive retention of the labor force in agriculture that cross-sector analysis identify (Gollin et al. 2014). While not all labor is unproductive in agricultural tasks, the ones that could be done with machines, or automated, are potentially the ones that drag the aggregated labor added value downwards the most.

Although a market constraint does not necessarily imply a market failure, the allocation of labor in such conditions could be inefficient under many circumstances. Initially, if there is a hoarding of labor, concentrating the labor force in a reduced set of farms, aggregated agricultural output would be lower compared to the one with a homogeneous distribution of this factor across the sector, due to the diminishing marginal returns of labor. Furthermore, if there is imperfect information, labor allocations would be biased in a way that the market would not be competitive. Unless it is assumed that both situations are not present in the market, the constraints in the supply of workers would be strongly linked to a labor market failure.

Proposition 2: *In general, labor market failures are positively correlated with household labor supply in agricultural activities.*

Proposition 3: *Adjustments in farms' demands due to a single market failure make them more prone to have multiple misallocated factors.*

3.3 Credit market failure

As it was explicitly exposed in section 2.3, the capacity for hiring, or investing, equilibrium levels of labor or machines fully depends on the capacity of farms for taking credits in financial markets. The availability and distribution of liquidity are then vital for overall productivity in the sector. Dividing the credit markets for both stages further generates tension on this part of the model, as the equilibrium allocation implies that credit supply and costs have to respond to the different needs of liquidity generated by tasks' heterogeneous costs. A failure on the credit market ($q_i^* > q_i$) would limit the capacity of farms to hire external factors (workers or machines) and, hence, effective market demands would behave similarly to the previous market failure cases, depending on the production stage when credit is insufficient, where leaning on the wage-free household members labor to complete equilibrium factor levels would once again help cover for market failures.

As the model is quite simple, the ways the credit market can fail are pretty straightforward. The demand side of the market can face problems if, given their characteristics (X_i), cannot afford to have equilibrium q_i^* at any stage. Although an important part of X_i is cultivated land, which keeps a strong correlation between demand and repayment capacity (as land is an ideal collateral for debt), discrimination in the system or systematic lending patterns can lead to liquidity constrain in some specific farms, affecting their factor provisioning. Take for example the case of a starting small farm that wants to invest in machines. If the price of machines is relatively high compared to the land, taken as collateral, the credit profile of these producers might not be good enough to have such levels of debt and lenders would not borrow enough money, creating the previously exposed threshold of investment on section 3.2 that would have profound effects on agricultural production. Another way to have the same problems under this framework could be the heterogeneity of interest rates (r). If, instead of assessing the capacity for repayment, credit lenders change the interest rate to cover themselves from the risk of no repayment, the real value of credits would be lower when a firm has a higher perceived probability of no repayment, implying, once again, the systematic differentiation of liquidity based on individual characteristics.

On the supply side, $\bar{Q_H}$ and $\bar{Q_{NH}}$ would be different based on production cost, although, ideally, both should be always at least equal to the demanded credit ($\sum_{i=1}^{N} q_i \leq \bar{Q}$). In this side of the market, overall lending capacity could be constrained, and, even if credit profiles are good enough for equilibrium credit levels, the limitations for having liquidity could lead to a systematic failure of production, with lower investment, factor demand, and, ultimately, a productivity gap. A low number of credit lenders or an extremely high cost of operations could lead to this situation of insufficient supply, however, is quite unfeasible taking into account the way informal markets adapt in rural areas (Giné 2011).

In short, to attain equilibrium production levels, credit markets have to be complete and competitive. Any other pattern in the lending process or inefficiencies when assigning credit could lead to individual choices of production below feasible levels and could potentially widen even more the agricultural productivity gap. X_i also implies no separability from productive performance and household characteristics when the constraints in credit bind. As credit profiles are strongly tied to the household's composition, it should be expected some degree of heterogeneity in this market, implying that positive discrimination could be needed in order to take the agricultural sector as a whole to its full potential. If X_i, in general, implies good enough credit profiles and supply is positively directed in order to ensure access, the credit market could grant the needed liquidity for all the producers and thus the farm performance would not be correlated by household characteristics.

Proposition 4: *Credit markets have to be complete and competitive in order to attain equilibrium production outcomes in agriculture*

3.4 Factor misallocation

Lastly, if we relax the assumption of cost efficiency on the most productive factor per task, defined as the ideal price level assumption (Acemoglu and Restrepo 2018), farms can have incentives to allocate factors in a task different from the efficient one. Firstly, rejection of ideal prices $(R + \Omega \geq \omega_H \frac{\beta^*}{\alpha})$ would lead to no investment in the Non-Harvest stage, that, when endowments are below equilibrium levels, would imply the following factor demands .

$$\hat{\mathrm{L}}_{NH} = \frac{\bar{T}_j}{\alpha}\frac{(1-\eta_{NH})}{\eta_{NH}} - \frac{\beta^*}{\alpha}K_i^0 \quad ; \quad \hat{K}_{NH} = K_i^0$$

Furthermore, misallocation could also be present in the harvest stage in this scenario if the cost of machine usage is low enough compared to relative wages. To put it more formally, there would be misallocation in harvest season if the following condition holds.

$$\Omega < \omega_H \frac{\beta}{\alpha^*} \tag{10}$$

This way, both Harvest and Non-Harvest tasks would use labor and machines in the production process as factor demands for harvest season would then be as follows.

$$\hat{\mathrm{L}}_{H} = \frac{\bar{T}_j}{\alpha^*}\frac{(1-\eta_{H})}{\eta_{H}} - \frac{\beta}{\alpha^*}K_i^0 \quad ; \quad \hat{K}_{H} = K_i^0$$

Both conditions could be solved for Ω and be simply stated in a two-way condition for this variable as follows.

$$\omega_{NH}\frac{\beta^*}{\alpha} - R \leq \Omega < \omega_H \frac{\beta}{\alpha^*} \tag{11}$$

In this case, although farms minimize their cost, factors are not properly allocated because of price levels, cost of factor usage, and relative productivity of factors.

Proposition 5. *If prices are not ideal, factor allocation is distorted by cost minimization choices that do not respond to relative productivities.*

4 Task-based production and coffee farms production

Once understood the basic assumptions, equilibrium, and potential productivity gaps, is important to inquire about the fitness of the model in the empirical study case environment. The central point to take into account when assessing the validity should be the central dynamic of production: Production separated by tasks and the comparative advantage of factors at different stages.

Due to the geographical characteristics of the region, most of the coffee plants are sown on land with a significant slope degree, making it difficult to automatize the harvesting process. Although Cenicafe, the Colombian R&D center for coffee, has developed technology that helps the recollection task, like field canvas and selective knockdown machines, hand picking is still the predominant practice. In the best-case scenario, most of those developments simply boost labor productivity in the task (α^*) and in no way is a substitute for the other factor. On the other hand, Cenicafe has developed plenty of machinery that facilitates the processing of coffee beans into commercial products (Oliveros et al. 2014). Cherry coffee beans, the raw output of coffee plants, can be processed in many ways, all the way up to toasted coffee beans ready for consumption. The most common commercial product in the region is dry parchment coffee and wet parchment coffee, which need a considerable amount of industrial processing, especially the first one. The main parts of this transformation process (peeling the pulp, washing and drying the beans) require machinery.

Empirically, the comparative advantage of factors would imply that farms that allocate labor in non-harvest tasks would be less productive than the ones that don't. Based on this premise, table 1 shows an OLS model for multiple productivity measurements, estimated over the same dataset, that explains the efficiency of production with various farm characteristics. The productivity measurements have different assumptions and, together, give a robust estimation of performance on the sample (Garcés Rodríguez and Gomez 2022). Throughout the four different measurements of productivity, which have pronouncedly different assumptions, having a positive level of working hours during the Non-Harvest season is negatively and significantly correlated with productivity levels.

While the way output is theoretically related to inputs varies across the methodologies, productivity terms already account for the role of aggregated labor in production. The scale effects of production, evident in the significance of aggregated labor and farm size, rule out the effect of having more productivity due to more factors. Therefore, the Non-harvest labor dummy would capture the existence of a dual status in labor usage, implying that this allocation of labor is negatively correlated with performance, as opposed to a farm that only places labor in the harvest season. Standard productivity analysis would simply aggregate factor levels and these results would be extremely counterintuitive, as a factor status should not affect negatively production under general substitutability assumptions. However, under the task-based production lens, the results are quite compelling.

Table 1: Farms characteristics associated with Productivity levels

	(1) DEA	(2) Fare-Primont	(3) Cobb-Douglas	(4) SFA
Having credit	0.1394	0.13467	-0.00649	0.04267
	(0.07479)	(0.08510)	(0.09672)	(0.09452)
Having Non-Harvest labor	-0.2289**	-0.27126*	-0.26783**	-0.39585***
	(0.08345)	(0.10931)	(0.09404)	(0.10258)
Distance to urban areas (Hrs)	-0.00004	-0.00078	0.00040	0.00008
	(0.00087)	(0.00090)	(0.00062)	(0.00096)
Female HHH	-0.1265	-0.19653	-0.27056*	-0.53141***
	(0.09326)	(0.11707)	(0.12301)	(0.12159)
Labor (Days)	0.00030***	0.00002**	0.00003	0.00009*
	(0.00007)	(0.00009)	(0.00003)	(0.00003)
Total machinery (COP)	0.01095***	-0.00266	-0.00379	-0.00337
	(0.00443)	(0.00335)	(0.00219)	(0.00315)
Total farm size (Hc)	0.00024***	0.00029***	0.00009*	0.00012**
	(0.00004)	(0.00006)	(0.00003)	(0.00003)
Share of farm area cultivated	-0.02897	0.2889	0.00253	-0.03778
	(0.1769)	(0.23018)	(0.20605)	(0.22176)
Property of the farm	0.7961	0.03584	-0.08555	-0.04167
	(0.1479)	(0.13869)	(0.1738)	(0.18632)
Municipality FE	Yes	Yes	Yes	Yes
R-squared	0.4213	0.2868	0.1243	0.09435
N	525	525	525	525

Note: Estimation of productivity variables are explained in depth in Garcés & Gomez, 2022. All numbers in parentheses are clustered standard errors at the municipality level.***,**, and * represent statistical significance at 0.1%, 1%, and 5%, respectively.

The rest of the elements exposed in the model are deeply connected with the empirical strategy, hence, the next section will extensively cover the feasibility of the remaining parts.

5 Empirical analysis

Although Propositions 1 and 5 are hard to test empirically and require a lot of detailed data over time, propositions 2, 3, and 4 are relatively more simple to analyze. The way farms demand and use factors in production can give valuable information about the functioning of markets. Testing for market failures is crucial to determine if the way production factors are distributed across farms is efficient and, therefore, are properly allocated in agriculture. While the empirical strategy does not directly estimate productivity, due to the complexity of estimating factor shares at each stage ($\alpha(j)$ and $\beta(j)$), the market analysis can determine if agricultural production is under ideal conditions.

In order to assess labor, machines, and credit markets' performance, three methodologies will be proposed. Firstly, for the labor market, a recursion test, as exposed by Dillon and Barrett (2017), will allow capturing heterogeneous access to hired labor. Secondly, for the capital market, a double hurdle model will be used in order to understand the capital accumulation process and decisions. Lastly, a Logit model will be used in order to understand the main determinants of access to the credit market.

5.1 Data

The dataset comes from a cross-sectional face-to-face survey of 654 coffee-producer households in 29 municipalities of Huila and Tolima, Colombia, realized by Federación Nacional de Cafeteros and

Alianza EFI. The survey included various modules on household labor, agricultural production, and financial behavior. For the production module, the high level of detail and disaggregation allow us to have a precise characterization of labor and machines used in agricultural production. Besides, data on land, construction, fertilizers usage, and technical assistance at the household level gives a complete picture of the farm production process. Other external datasets for prices (SIPSA) and weather and soil quality were used to complement the survey data.

Labor data was disaggregated by periodicity (seasonal or regular) and by origin (hired or unpaid family members), which perfectly allows us to analyze the different production stages. Machines are the subset of the capital captured in the survey, around 10 classes of industrial, semi-industrial, and crafted tools, that are only used in the Non-Harvest production stage, namely, the ones that do the processing of the cherry coffee beans. Households' socioeconomic characteristics are disaggregated at the individual level, which allows a detailed characterization of each individual. Lastly, a module of financial inclusion captured the usage and request of multiple formal, semi-formal, and informal credit instruments taken in the previous 12 months.

Table 2: Summary statistics

Statistic	N	Mean	St. Dev.	Min	Max
Cultivated area	525	63,008	99,994	1.500	1,116,000
Machines (COP)	525	5,089,395	3,464,872	0.000	38,823,241
Total farm area	525	92,299	224,444	2,250	4,120,000
Labor	525	7,334.141	11,755.020	24	134,400
Prime male share	525	0.368	0.252	0.000	1.000
Prime female share	525	0.305	0.215	0.000	1.000
Elderly male share	525	0.148	0.240	0.000	1.000
Elderly female share	525	0.101	0.202	0.000	1.000
Total Household members	525	3.019	1.303	1	9
Average education (Years)	525	6.322	2.965	0.000	16.000
Distances to urban areas (minutes)	525	57.747	53.059	0	360

5.2 Labor market

At both harvest and non-harvest stages, farms take choices over hired labor and household members' labor allocation. As the discussion in section 2.2 hinted, if the agricultural labor market is complete and competitive, there will be a clearing market wage in this market, and household members would take labor allocation choices based on this wage, as it will correspond to the marginal benefit of working in the farm. More importantly, the proper functioning of markets would make indistinguishable the contribution of hired labor and family labor as agricultural aggregated demand would be satisfied and household members, if choosing to work in the sector, could do it in any other farm for the same benefits. In general, complete and competitive markets can efficiently satisfy factor demands in a way that households do not depend on their endowments for production, therefore, factor levels would not be correlated with household characteristics.

In the task-based model, optimal levels of factors a determined by the task elasticity of the factor (η), the exogenous and fixed cultivated land ($\bar{T}_i$), and credit availability (q_i). To get to those demands, condition 5 must hold, so they also depend on prices, wages, and relative productivity. Lastly, in the Non-Harvest stage, the demand for labor would also depend on the existing stock capital (K_i^0).If we assume that $q_i = q_i^*$, all these previous elements are exogenously determined, and thus we could define equilibrium demands as a function of these variables as

$$\mathrm{L}_H^* = L_H(\omega_H, R, \Omega, \alpha^*, \beta, \eta_H, \bar{T}_i)$$
$$\mathrm{L}_{NH}^* = L_{NH}(\omega_{NH}, R, \Omega, \alpha, \beta^*, \eta_{NH}, \bar{T}_i, K_i^0)$$

This implies that, under these settings, households solve the profit maximization problem based on exogenous variables. Hence, equilibrium production and utility choices would be taken separately and recursively, as the variables that determine utility are unrelated to the production ones. This is known as the separability hypothesis and was originally proposed by Benjamin (1992). It can be easily tested with an OLS model by regressing labor used in production with variables that affect consumer behavior but do not affect factor levels, basically, household characteristics (Z_i), known as the recursion test.

As Dillon & Barrett argue (2017), the main characteristics that determine consumption and, if the assumption hold, should not be related to factor demands are household demographic characteristics, like total members, age, and gender. As these are also the main determinants of the household labor force, the test can effectively capture if the labor market allocates labor efficiently enough such that households do not need to rely on their own labor endowments. Although the test is focused on the labor market, rejecting the separability assumption does not imply that such a market is the source of the problem. Instead, it shows that at least one of the markets is incomplete, due to the substitution effect that households would incur if one of the markets fail.

We will follow the methodology proposed in Dillon & Barrett (2017) as they also test the recursion hypothesis with cross-sectional data. Household, sex, and age are clustered into four groups expressed as a share of the overall household members: prime male share, prime female share, elderly male, and elderly female, where prime is between 10 and 60 years and elderly is above 60, while total household members are defined as the number of members that are at least 10 years old (minimum working age in Colombia rural areas). Apart from the household's characteristics, the model controls for the other factor variable that theoretically determines labor scale, such as land cultivated, and machine levels. Other control variables that can potentially imply differences in the demand for labor will be included: A dummy variable for producing certified coffee, as it may require special practices, sex of the head of household, distance to the urban area in minutes, to control for the ease of getting workers, average years of schooling in the household, that implies different ω^M and ultimately, different available workforce, and farms antiquity to account for the expertise and stage of development of production. Multicollinearity in the control variables was rejected and thus is included altogether in the model. Table 11 in the appendix displays the variance inflation factor of the OLS model, where the multicollinearity rejection is tested.

Lastly, in order to capture the financial capabilities of the households and the relevance of credit in the production model, separability will be tested within credit status (Having or not having credit for agricultural activities), accounting for the heterogeneity caused by different financial constraints. The specification of the test, in this case, would be as follows

$$ihsL_i(j) = \alpha_0 + \beta_1 ihsT_i + \beta_2 ihsK_i(j) + \beta_3 ihsZ_i \times Q_i + \beta_8 C_i \times Q_i + \epsilon_i \tag{12}$$

Where j $\in$ [Harvest, Non-Harvest], ihs is the inverse hyperbolic sine transformation, L_i^H is total labor hours used in the task j, T_i is total land cultivated, K_i^H is the price of machines used in production, Z_i are the characteristic variables (sex and age share groups and total household members), Q_i is credit status, C_I are control variables and ϵ_i is the error term, clustered at the municipality level. Multiple specifications were used in order to test the homogeneous credit market assumption. Appendix 8 has the results of equation 12 with multiple credit markets (formal, informal, and mixed credit markets), however, the main specification results hold. This way, the separability hypothesis would be rejected if β_3 is statistically different from 0 (Ho: $\beta_3 = 0$).

Table 3: Regression results for the OLS recursion test

	Harvest labor			Non-Harvest labor		
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable:						
Total labor hours per task						
Total cultivated land	0.59648***	0.48818**	0.48597**	0.19268	0.1626	0.15799
	(0.17754)	(0.17182)	(0.17556)	(0.20389)	(0.2198)	(0.21527)
Machines	0.08757	0.07700	0.07011	0.06822	0.08838	0.08965
	(0.08066)	(0.07021)	(0.07198)	(0.24643)	(0.25209)	(0.24397)
HHH sex		0.07302	-0.28889		-0.75865	-0.41097
		(0.33427)	(0.21844)		(0.80771)	(0.59298)
Distance to urban area		-0.00253	-0.00093		-0.01293	-0.00934*
		(0.00282)	(0.001715)		(0.00793)	(0.00472)
Certified Coffee		0.28353	0.32393		2.9724**	0.69450
		(0.30830)	(0.27056)		(1.0322)	(0.69878)
Average education		0.13449**	0.09841**		0.0000	0.10307
		(0.04799)	(0.03163)		(0.1406)	(0.08583)
Farms antiquity		0.00500	0.00016		-0.05471	-0.0367
		(0.01057)	(0.00778)		(0.02813)	(0.02124)
Households with credit						
Household members	-0.02342	-0.05756	-0.1055	0.61191	0.7091	0.76000
	(0.23543)	(0.25523)	(0.24924)	(0.6853)	(0.7619)	(0.74929)
Prime Female share	1.43538	0.65988	0.39244	1.9446	1.0367	1.36079
	(1.01028)	(1.03618)	(1.02247)	(2.56715)	(2.7696)	(2.67416)
Prime male share	1.99274*	0.69414	0.6833	4.56464	3.3442	3.48985
	(0.82048)	(0.86991)	(0.86163)	(2.47642)	(2.7043)	(2.61300)
Elderly Female share	0.84670	0.65216	0.22156	1.8139	2.8010	3.1021
	(1.0251)	(1.11466)	(1.03895)	(2.5526)	(3.0048)	(2.8793)
Elderly Male share	0.92851	-0.23475	-0.28236	2.62211	2.1218	2.48311
	(1.03925)	(0.17155)	(1.10039)	(2.54834)	(2.8454)	(2.71607)
Households without credit						
Household members	-0.1992	-0.26815	-0.20631	2.02656**	2.2637**	1.98756*
	(0.3031)	(0.33750)	(0.33098)	(0.75164)	(0.86056)	(0.82998)
Prime Female share	2.1115	0.55512	1.03109	0.75419	3.4300	2.02609
	(1.21216)	(1.26848)	(1.21486)	(2.74709)	(3.1287)	(2.89953)
Prime male share	1.99037	0.40745	0.59429	-0.73673	-2.3104	-2.75291
	(1.10505)	(1.17757)	(1.12278)	(2.68247)	(3.0796)	(2.88828)
Elderly Female share	0.76277	-0.00404	0.46673	-0.64373	0.97185	-0.08185
	(1.12825)	(1.12155)	(1.10957)	(2.67194)	(2.8779)	(2.73162)
Elderly Male share	0.28579	-0.57607	-0.52668	0.90908	2.3235	1.45187
	(1.03754)	(1.13994)	(1.08555)	(2.59494)	(2.9129)	(2.77627)
Municipality FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls in credit status	No	Yes	No	No	Yes	No
R-squared	0.3266	0.3531	0.3436	0.1956	0.2393	0.2297
N	527	527	527	527	527	527

Note: All numbers in parentheses are clustered standard errors at the municipality level. All production factors are inverse hyperbolic sine transformed. When control variables interact with credit status, the results in the table are the interactions with no credit. ***,**, and * represent statistical significance at 0.1%, 1%, and 5%, respectively.

Table 3 shows the results from equations 12 regression for harvest and non-harvest labor respectively. Columns (1) and (4) show results without control variables, columns (2) and (5) are the results when interacting controls with credit status, and columns (3) and (6) are the results when controls are included as factor determinants. As expected, the land is a main determinant of the scale for labor in harvesting activities, given the high elasticity of space and coverage per worker, estimated at 0.48 by the model. On the other hand, the cultivated area is not determinant for non-harvest activities, showing the different nature of the process. These results could hint at some patterns in the parameters of the model and thus could be expected that the share of land in harvest activities is greater than in non-harvest ($\eta_H > \eta_{NH}$). Finally, average years of education are positively correlated with labor demand for the harvest season, where one extra year of education for the household, on average, is associated with an increase of 9.8% in the demand for harvest labor.

The recursion hypothesis is not rejected for households with credit. For both harvest and non-harvest stages, labor demand is not related to any household's characteristics, and thus, it could be argued that these farms behave like they face competitive and complete markets. However, recursion is rejected for farms without credit, as there is a positive and significant correlation between Non-Harvest labor demand and total household members for these observations. Credit role is, thus, proven to be a determinant in agriculture production, as having, or not having, liquidity defines the way farms behave. Once again, rejecting separability does not imply that the Non-Harvest market is failing for households without credit, but that, at least, one of the factor markets is. However, the allocation behavior is quite telling: When there is no credit, having an additional household member is associated with allocating about twice as much non-harvest labor.

What would be the logic of this behavior under the task-based model? First, positive labor demand at this stage implies a misallocation of the factor, that could be caused by deficient machine levels. Second, the elasticity of household members and the Non-Harvest labor scale implies an important degree of depth in the market failures, as, in this setting, the farms mostly end up relying on their own endowments to provide labor to the task. Third, it could be argued that, in credit-less farms where the recursion test is rejected, this behavior could be caused by insufficient liquidity that leads to no investment levels and the inability to hire workers. In short, although misallocation is evidenced, and these choices are associated with behaving as if there are no complete or competitive markets, is not clear what could be causing these patterns in allocation.

As the previous analysis shows, rejecting separability raises more questions than answers, and any meaningful conclusion requires the analysis of other markets. However, it is important to take into account that labor takes up to 64% of total production costs in Colombian coffee farms (Serna et al. 2010), therefore, any issue in this market almost certainly implies production problems. The pivotal role of labor cost in the budget pushes producers to be efficient in this allocation and thus it would not be feasible that, under the light of these findings, the problem is completely outside this market. Up to this point, however, it can only be asserted that farms without credit do not behave as markets were complete and competitive.

Before getting into other markets to understand what is happening, labor market efficiency can be approached in two other ways. First, we can analyze household labor allocation determinants to understand more the rationality behind these choices. Secondly, and testing one of the assumptions of the model, we can analyze the potential endogeneity of land and household labor in order to understand how labor endowments determine production choices.

As household members seem to be an important proportion of Non-harvest labor in credit-less settings, understanding the way this group of people allocates their labor force could help us to understand the recursion test results. A simple logit model for monthly labor sector allocation (working on the farm or in the non-agricultural sector) at an individual level, separating for credit status in the farm, can shed some light on overall market dynamics for household members in both types of settings. Given the results in table 3, it would be expected that the way household members allocate labor would differ based on the credit status. More specifically, if there are labor market failures, we could envision that households without credit retain their members in agricultural activities in a higher proportion than households with credit, as the model predicts.

Table 4 shows the results of a Logit and Probit model where Y = 1 if a household member at month t chooses to work outside the farm and Y = 0 if chooses to work in the farm. The model takes a subset of the sample of the months where people effectively worked, uses as explanatory variables a set of personal and household characteristics, and, Lastly, includes a dummy for the harvest season to account for the heterogeneity in labor needs in farms. For both groups, education years and being a female are strongly associated with a higher probability of working outside the farm, showing, once again, the relevance of human capital for labor choices, as well as the gender role dynamics in rural households, also evidenced in the significance of male share in the probability of working outside the farm for both groups.

Table 4: Monthly sector choices of labor allocation

	With credit		Without credit	
	(1)	(2)	(3)	(4)
	Logit	Probit	Logit	Probit
Dependent variable:				
Sector chosen for labor (=1 if Non agro)				
Harvest season (=1 if true)	-0.09279	-0.04455	-0.0593	-0.02697
	(0.29887)	(0.16278)	(0.4326)	(0.2548)
Years of education	0.48609***	0.26539***	0.41365***	0.20955***
	(0.03748)	(0.01968)	(0.04373)	(0.0179)
Sex (=1 if female)	2.46436***	1.32233***	2.34905***	1.22559***
	(0.2359)	(0.1220)	(0.26939)	(0.13378)
Machines	0.59144**	0.34855***	-0.95692**	-0.43168*
	(0.18049)	(0.09767)	(0.37049)	(0.20608)
Total cultivated land	-0.07977	-0.04601	-0.18692	-0.12941*
	(0.05181)	(0.02818)	(0.08985)	(0.05334)
Distance to urban area	0.00944***	0.00499***	-0.03092***	-0.01879***
	(0.00198)	(0.00096)	(0.00443)	(0.00256)
Certified Coffee	-0.77808	-0.16852	1.4348***	0.86391***
	(0.32183)	(0.18172)	(0.28565)	(0.17213)
Household members	-0.24012	-0.13485	0.5222	0.33764
	(0.24611)	(0.12971)	(0.40251)	(0.20769)
Prime Male Share	3.14602**	1.8842***	1.3036	0.14905
	(0.9884)	(0.54757)	(1.06703)	(0.5229)
Prime Female share	2.03243	1.26739	2.3700	0.79003
	(0.99977)	(0.50344)	(1.0111)	(0.50368)
Elderly Male share	0.49281	0.63014	6.2317***	2.71789***
	(1.3827)	(0.7829)	(1.3475)	(0.63222)
Municipality FE	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
AIC	968.06	964.9	653.9	659.7
N	4545	4545	3528	3528

Note: All numbers in parentheses are clustered standard errors at the household level. All production factors are inverse hyperbolic sine transformed. ***,**, and * represent statistical significance at 0.1%, 1%, and 5%, respectively.

Although being in harvest season is not determinant in sector choices, meaning that, on average, there are no on/off farm labor jumps throughout the productive process, the particular, and heterogeneous, relation between machines might be the key to understanding household labor behavior. In households with credit, an increase in the machine levels makes more likely the allocation of labor outside the farm, as expected due to substitutability, given that overall labor demand would decrease. However, when there is no credit, machines play the opposite role, as a marginal increase in the scale makes people less likely to work off-farm. The lack of credit, which heavily conditions the capacity to hire workers, might imply an important degree of reliance on labor endowments such that, in all seasons, production is mainly based on household labor, and a bigger scale in machines then implies a higher opportunity cost for leaving the farm. These results could suggest that the behavior of farms without credit, captured in the separability test, is as if there are no complete and competitive markets because family members misallocate their labor force in agriculture throughout the season and production scales due to liquidity constraints that retain them in their farms and not to labor market failures themselves.

While the previous results unfold the way the misallocation is related to credit access, it is unclear if the market failures are on the labor market, Harvest or Non-Harest, or not. Testing the land exogeneity hypothesis could be an ideal way to understand labor market efficiency given the high dependence of both factors. Literature has found that cropping area choices are heavily correlated with the total number of household members, as it is a determinant of available labor, and thus, the capacity of exploitation over the land (Eastwood et al. 2010), directly rejecting the efficiency of markets in agriculture to allocate workers. If, as the literature suggests, choices over cultivated land and household labor force are so intertwined, testing this hypothesis within both labor seasons could capture the relative role of markets in farm choices and, thus, potential labor market failures could be identified.

As it is assumed that hired labor and household members are perfect substitutes, the share of household labor in each task is a good initial indicator of current market status, as, at any scale, this type of labor shouldn't constitute an important share if markets are efficient. However, testing the correlation of this share with land choices could show a pervasive labor market dynamic that leads to resizing the exploitation based on a deficient labor supply. Given the middle/long-term nature of choices taken over the land, any endogeneity between land and household labor share would capture strong and constant dynamics in labor markets, not shocks or temporal fluctuations. Furthermore, it could also be problematic for the recursion test identification as household labor would be endogenous to cultivated land.

A first approximation to the problem shows that household characteristics are not related to the total hectares cultivated, addressing any concern over potential endogeneity on the recursion test (See appendix 8). However, the share of the cultivated area over the total farm area is negatively correlated with the number of household members(See appendix 8)implying that bigger households tend to exploit less area of their farms. Although it is counterintuitive to what has been found in the literature, especially as total farm size is also not significantly related to household characteristics (appendix 8), the significant correlation between labor endowments and production choices is a first indication of deficient labor markets.

Table 5 shows the results of the endogeneity test controlling by various household characteristics. Intuitively, the higher the share of family labor, the more dependent is production on household members, and therefore, the land scale would be downwards resized to adapt to this capacity of exploitation. Table 5 shows that the number of household members is positively associated with the share of household labor in agriculture, as it directly means a higher capacity to produce tasks. Furthermore, the total cultivated area is negatively correlated with overall family labor share, however, there is heterogeneity in the relationship of this variable within the different labor seasons. While for Non-harvest labor there is a strongly significant correlation with the share of household labor, where a marginal increase in cultivated area is associated with a decrease of 7% on the share of household labor, there is no significant relationship with harvest labor.

Given the predominance of labor in the harvest season, the no correlation between family share at this stage and cultivated land could be interpreted as the independence of scale exploitation and labor endowments, which, together with the recursion test results, make a strong case for the efficiency of the harvest labor market in allocating workers. Otherwise, the importance of labor in harvesting would lead to resizing of the production scale if the markets did not provide the needed labor. However, the scale and Non-Harvest labor correlation is quite ambiguous due to the role of labor in this task, as resizing the scale based on the shortage of an unproductive factor is not something that would be expected. The fact that a household is using labor at this stage is indicative of low capital endowments, which forces the demand for labor in a scenario where it wouldn't be expected to find a properly developed market, as it was argued in section 3. Furthermore, the restricted capital scenario could be heavily correlated with other factors that constrain the size of the exploitation of the farms. Altogether, a misallocation of labor within a constrained exploitation scale could be taking place in such farms due to initial inauspicious conditions that trap the farm into an unproductive cycle of production and lead to production choices that limit the development of the farm. Therefore, the results from table 5 could be interpreted as the results of what was theorized as a double productivity gap in section 3.2, where the constraints in other markets derive into the misallocation of labor in agriculture.

Table 5: Land and Household share endogeneity

	(1) Total Labor	(2) Harvest Labor	(3) Non-Harvest Labor
Dependent variable: Share of household labor on Agriculture activities			
Cultivated land	-0.05161*	-0.00537	-0.07094***
	(0.02052)	(0.00607)	(0.02095)
Distance to urban areas	-0.01175	-0.00271	0.01631
	(0.01914)	(0.00446)	(0.02692)
Average education	-0.00390	-0.00082	-0.00997
	(0.00550)	(0.00257)	(0.00805)
Have credit	-0.02157	-0.00517	-0.03396
	(0.03203)	(0.01326)	(0.04471)
Total HH members	0.15834***	0.02261	0.18222**
	(0.04025)	(0.01775)	(0.05733)
Prime Female share	0.00875	0.03742	0.05943
	(0.10443)	(0.05489)	(0.15129)
Prime Male share	-0.06415	0.01746	-0.12377
	(0.07816)	(0.02530)	(0.11828)
Elderly male share	0.04123	0.03508	-0.01561
	(0.11411)	(0.03581)	(0.18892)
Female HHH	-0.02105	-0.00272	-0.03605
	(0.04218)	(0.01901)	(0.05564)
Certified coffee	-0.02618	0.00845	-0.04935
	(0.04539)	(0.02183)	(0.06495)
Farms antiquity	-0.00239	-0.00006	-0.00061
	(0.00144)	(0.00061)	(0.00188)
Total machinery	-0.02444	-0.00324	-0.06184
	(0.02084)	(0.00832)	(0.03326)
Municipality FE	Yes	Yes	Yes
R-squared	0.2802	0.0953	0.4335
N	525	525	525

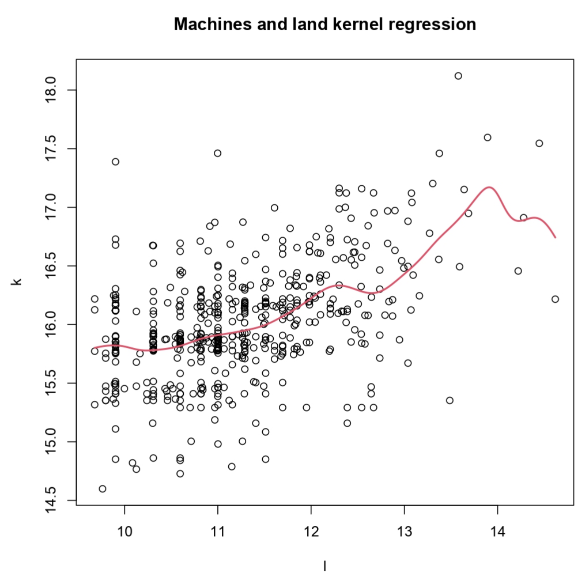
Note: All numbers in parentheses are clustered standard errors at the municipality level. Factor variables are inverse hyperbolic sine transformed. **,**, and * represent statistical significance at 0.1%, 1%, and 5%, respectively.

In conclusion, the labor sector analysis showed that propositions 2, 3, and 4 of the model were proven to be aligned with the empirical results. First, the recursion test showed the relevance of credit markets in equilibrium allocation that proposition 4 argues, as such farms behave as if markets were complete and competitive. Second, when the recursion test was rejected, households allocated their labor force in a higher proportion to agriculture as Proposition 2 argues. lastly, as proposition 3 argued, adjustments in allocation due to market failures make more likely the misallocation of other factors, exposed in the land exogeneity discussion. While the results are quite compelling, and completeness was not rejected for the harvest labor market, these conditions are not enough to assure auspicious production outcomes. More specifically, machines and credit markets play a crucial role in the production process and, therefore, should be analyzed and contrasted with these results.

5.3 Machine market

Given the role that the Non-Harvest labor market played in the recursion test and the land exogeneity analysis, and the fact that labor should not be allocated at this stage of production, is important to understand machine market dynamics and the allocation of this factor across the sample. More specifically, is key to understand how households acquire new capital and, in order to test misallocation, check for machinery distribution across farms. Analyzing a long-term process such as capital accumulation with cross-sectional data is quite limiting. However, the data set is rich enough to allow us to have some meaningful insights. The survey had a machinery section that covers the tenure of a group of common machines, both semi-industrial and handmade, capturing the number of units and the year they were bought. Ideally, one would measure capital accumulation dynamics through time taking into account labor market dynamics and liquidity constraints at each moment of time, unfortunately, past labor and credit data are not available. Still, the latest machinery acquisitions can be analyzed and it could derive some insights from this market.

An initial analysis of aggregated machine levels hints that there is an efficient distribution of the factor across farms. No household characteristic is correlated with stock capital and the only determinant of this variable is cultivated land (see appendix 8). The simple kernel regression in the graph 5.3 could be indicative of an overall efficient allocation of capital, as land should heavily be correlated with machines in ideal settings. However, this does not imply efficient markets or ideal scales for production, it simply states the expected relationship between cultivated land and machines.



In order to analyze the latest acquisition of the machines (Investment at the moment of the survey), a double hurdle model will be proposed. As originally proposed by Cragg(1971), the methodology is quite similar to the Tobit models, dealing with corner observations within a distribution. As decisions over investments are not taken every period of time, the methodology has to deal with the censoring of observations, which makes it ideal for this type of analysis (Martínez-Espiñeira 2006; Simtowe and Zeller 2006). Intuitively, the model divides consumption decisions into two stages: first, households decide whether or not to access the machines market, and then, if they decide to enter, take a separate decision of how much of the good should be bought. The first hurdle would be defined as follows

$$d_i = 1 \text{ if } d_i^* > 0 \text{ and } 0 \text{ if } d_i^* \leq 0$$

Where

$$d_i^* = z_i' \, \beta + \epsilon_i$$

Where d_i^* is the latent decision to enter the market of new machines and d_i is the actual choice. Once decided whether or not to enter the market, the second hurdle could be defined as follows

$$y_i = y_i^* \text{ if } y_i^* > 0 \text{ and if } d_i^* > 0.$$

$$y_i = 0 \text{ Otherwise.}$$

where

$$y_i^* = x_i' \, \alpha + u_i$$

Where y_i is the observed quantities of machines bought. Overall specifications of the model will depend on the relationship and dominance of both hurdles, where the orthogonality of ϵ_i and u_i would determine if the model is independent or dependent. Lastly, the model is estimated by maximizing a log-likelihood function that will be defined based on the independence of the error terms (Wodjao 2020).

Following Simtowe & Zeller (2006) A simple double-hurdle model will be proposed, where a single model will be used to specify both hurdles. Once again, household characteristics and credit status will be included in order to account for market efficiency (in the same way that the recursion test does) and liquidity constraints, respectively. In addition, the previous machine scale (denoted as K_i^0 in the model), to account for investment feasibility, farms' antiquity, to account for expertise and stage of farm development, and technical assistance will be included in the model. Table 6 shows the results of the double hurdle model for the machines bought in the year that the survey was realized.

Table 6: Double hurdle model for machinery market

	First Hurdle	Second Hurdle
Dependent variable: **ihs I**		
ihs farms antiquity	-0.0341	0.0006
	(0.0884)	(0.0042)
ihs K_0	-0.1049***	-0.0048***
	(0.0157)	(0.0006)
Technical assistance	0.351**	0.0138*
	(0.1233)	(0.0058)
Having credit	-0.0160	-0.0033
	(0.1182)	(0.0056)
ihs Household size	-0.0134	0.0035
	(0.1584)	(0.0077)
Prime male share	0.4542	0.0277
	(0.3992)	(0.0195)
Prime female share	0.8252*	0.0535**
	(0.3704)	(0.0180)
Elderly male share	0.4298	0.0389
	(0.4659)	(0.0236)
Municipality FE	YES	
Coefficient of determination	-15243	
Likelihood ratio index	-2.2275	

The results of the Double hurdle model show the importance of the role of K_i^0 (endowments, in the model) in machine choices. As expected, the higher the endowments, the fewer new machines are needed in order to get to equilibrium levels, thus, K_i^0 has negative and significant effects on the likelihood of both accessing the market and having positive investment levels. Although these results are contradictory to condition 6, where, the more endowments, the more likely to invest, the latter should be interpreted more as a threshold for having incentives for investment, due to the fact that the corner solution in investment arises from the static nature of the model. Technical assistance, on the other hand, happens to be a positive public policy in terms of productive transformation, as the recipients of this program are more likely to have positive levels of investment. This program is sponsored by Federacion Nacional de Cafeteros and, exogenously, assigns a technical professional to a region that visits the farms in order to give technical advice.

Very interestingly, while overall machine levels are uncorrelated to household characteristics, the double hurdle model rejects this behavior for the latest investments in machines, as prime females' share is positive and significant in both hurdles. These findings could imply that the machines' accumulation process was inefficient, to begin with, as these new biased investments add to a balanced capital tenancy on the aggregated level in a way that seems to adjust previous imbalances. It is also quite fascinating the household characteristics that cause the bias if we recall table 3.2 results, as females were more likely to work off-farm in general. A household with a higher share of females would then be more likely to face a limited number of the available workforce for agricultural activities. Although the literature has interpreted the relation between capital investment and off-farm labor through the ease of investment, given that off-farm wages tend to be higher (Simtowe and Zeller 2006), under the task-based model, it could also be argued that, in general, farms would have more incentives for investment when the substitute (labor in this case) of the factor is hard to hire, that could be due to market failures or high rate of off-farm household labor, or both.

While we cannot assert causality given the limitations of the dataset, this machine accumulation bias should be addressed in future research and could clarify the relationship between machines and the household labor force. Still, in light of the double hurdle model results, it can be said that machine accumulation is biased and are not in line with the equilibrium results, as it is correlated with household endowments, hinting, once again, at the presence of incomplete markets in the sample.

5.4 Credit market

Throughout the theoretical and empirical analysis, credit has been highlighted as a crucial element in agricultural production choices. Both the recursion test results and the theoretical model show that the credit market must be complete and competitive in order to have equilibrium production levels. In order to test this condition, however, is required a time series of the sample, and therefore, the analysis is limited by data constraints.

The agricultural credit market on its own deserves an in-depth study, but, given the limitations of the sample and the way the market is modeled, a pretty straightforward question could give important results: Does everyone have the same probability of accessing financial instruments? If there are patterns or access barriers that create involuntary segregation from financial markets, excluded households would face harsher liquidity constraints and, as non-harvest labor market analysis showed, ultimately would derive in labor market failures.

The theorized connection of factors provisioning and credit markets, exposed in section 2.3, is far more complex than what could possibly be derived from analyzing credit market access, but, in this context, other elements are either not present or insufficient to understand the phenomenon. First, as it will be proven later, credit quantities appear to be irrelevant to the functioning of labor markets. Second, as it was found on the recursion test, accessing credit markets, whether formal, informal, or both, is enough for producers in order to provide themselves with factors in a way that seems like they face competitive markets. From this finding, it is also argued that, while interesting, there is no necessity for comparing formal with informal lenders' characteristics. Third, credit acceptance is not an issue in the sample. Credit data is reported from a set of questions that captures every type of credit request and whether or not the credit was ultimately taken or not. In the sample, less than 3% total credit requests were rejected, showing that, whenever households want credits, they almost certainly end up getting them. While, also because of the limitations, no conclusions can be derived from these results, these findings allow us to rule out these elements from the analysis.

Why, then, would asking for credit be a relevant variable? if households can have credit whenever they ask for it, this variable would only capture the need for credit at the evaluated point. However, the findings of the recursion test depend on this status, and any heterogeneity in credit access would then imply that the results from the labor section must be contrasted with this distribution of the credit market. In order to assess this question, a Logit model will test the determinants of accessing the credit market. In order to define the variables that compose X_i, we will follow the specification proposed by Coy (2017), which studies the determinants of credit access in rural Colombia.

Table 7 shows the results for the logit model and a double hurdle model for credit request and credit values, respectively. The logit model and the first hurdle, by construction, show the determinants for requesting credit. Head of household years of education is positively and significantly associated with the probability of asking for credit while having a female head of household is associated with a decrease of 12% in this variable. This means that, for the recursion test, while educated households are more likely to have credit, households led by females are far more likely to be on the side of the economy that behaves like there are no complete and competitive markets, due to this lacking credit. The heterogeneity on this side of the market means that the productivity gap is more common in certain types of households merely based on their characteristics, not their production choices.

The second hurdle, which captures the choices over quantities of credit, has some interesting results as well. As scales of production requirements, cultivated land areas are significantly correlated with the amount of credit taken, a positive sign of the assignation of credit in the market. Being a recipient of the technical assistance program is also associated with a higher amount of credit, that, given the fact that is exogenously determined, would not directly imply a key characteristic that allows higher debt levels. Moreover, this positive correlation found between technical assistance and credit size could also be derived by the correlation of the former with the propensity for machinery investment, which would lead these households to take bigger credits.

Table 7: Credit market access and demand

	Logit Model	Double Hurdle	
		First Hurdle	Second Hurdle
Age	0.00442	0.01578 (0.02861)	0.0013 (0.0025)
Age squared	-0.00005	-0.00023 (0.00026)	-0.00001 (0.00002)
Education	0.01458*	0.03724* (0.01853)	-0.00127 (0.00157)
Property over the farm	0.03068	0.06481 (0.13823)	-0.00067 (0.01175)
Technical assitance	0.03418	0.13378 (0.12923)	0.02441* (0.01097)
Distance to the bank	0.00075	0.00201 (0.00153)	-0.00000 (0.00001)
Cultivated land	0.00000	0.06196 (0.05705)	0.01749*** (0.00504)
Side crops	0.02509	0.04740 (0.12564)	-0.0088 (0.00103)
Household members	0.01727	0.26154 (0.13803)	0.00298 (0.01219)
Female head of household	-0.1246*	-0.3229* (0.15408)	0.01166 (0.01430)
Cooperative member	0.04325	0.07481 (0.13495)	0.00676 (0.01096)
Municipality FE	Yes	Yes	Yes
N	508	529	529

Note: ***, **, and * represent statistical significance at 0.1%, 1%, and 5%, respectively.

Although these results allow us to identify heterogeneity in the composition of productive and unproductive farms, we cannot assert any other pattern or inefficiency in the credit market. While the logit model results could hint at a self-selection for female-led households, these types of households could have faced differentiated credit situations in the previous production cycle that led to a different behavior at the evaluated period. Still, the heterogeneous composition of credit status does imply an uneven representation of households led by women in the unproductive side of the economy.

6 Discussion

The task base model has allowed us to guide the analysis and understand factor allocation behaviors, especially labor demand and sector allocation, that otherwise could have been quite puzzling. Separating market analysis by season was crucial for the recursion test and allowed us to capture market failures, something that would have been rejected if the test was realized with aggregated labor demands (see table 12 in appendix). Moreover, the land homogeneity test also allowed us to not reject the completeness of the harvest labor market precisely because of this separate approach. Besides the discussion in section 4, that argued a proper fit of the model to this productive setting, propositions

2, 3, and 4 also seem to be in line with the empirical results, proving, once again, the advantage of taking the task-based approach to analyze coffee productivity.

The limited findings that current recursion test literature can address demand a more profound dive into the potential causes of market failures and the overall effects of the productivity gap. While the latter approaches to the test are an improvement, taking heterogeneous effects into the analysis (Merfeld 2021), there is still quite a lot of work to do in order to fully capture these market dynamics. This paper has highlighted the conceptual advantages of approaching factor allocation through a holistic analysis of multiple markets, that allows us to understand the substitutability dynamics incurred in production and have a broader picture of the factor demands, allowing us to identify more sharply the possible roots of market failures. However, it must acknowledge that is heavily demanding in terms of data availability and, as the empirical analysis limitations showed, even if the panel data of this survey was available, many things will be unable to be captured in order to properly assert the market failure causes, like initial endowments, for example.

Throughout the empirical analysis, systematic correlations of endowments and factor demands were found in the factor provisioning behavior, implying that inefficient markets are the pervasive settings in Colombian agricultural production. While it is still unclear the source of the failure, having such patterns in labor, machines, and credit markets could hint at an important level of the deepness of the problem. It is worth noting, however, that the no rejection of completeness in harvest labor markets is a positive finding for the prospects of coffee production in Colombia and the understanding of the productivity gap in the region. Once again, this finding was possible due to the theoretical approach to agricultural production, which makes a strong case for the revisiting of market definition and recursion test results under more desegregated lenses.

7 Conclusion

In this paper, we have approached the agricultural productivity gap through the lens of a task-based production function. The theoretical section provided a sequential production setting that pointed out previously unexplored causes of the agricultural productivity gap, highlighting the interdependence of factors demands. In the empirical section, multiple market analyses showed the pervasive market failures existing in rural Colombia that distorted the equilibrium demands in coffee production.

The conclusions can be summarized as follows. First, coffee farms in the sample demands of labor and capital are correlated with their household characteristics, implying that there are market failures in the sector that prevents the equilibrium allocation of factors across the economy. Second, the inefficient allocation of labor tends to be accompanied by land scale constraints, meaning that efficiency in one-factor allocation relies on the choices, or outcomes, of other factor levels and markets, and therefore, one market inefficiency can derive in a double productivity gap. Lastly, given the lending pattern behavior of credit, unproductive behaviors distribution is biased in the sample, as farms led by females are overly represented in the side of the economy without credit.

For future research, further analysis of factor provisioning, especially machines, and credit, are needed in order to properly identify the causes of the market failures and misallocation dynamics in this production setting. Moreover, given the adequate fit of the model, the task-based approach could be taken into other crops or production settings to have a novel understanding that could lead to new ways of understanding agricultural productivity. Lastly, the empirical analysis must be updated and refined when panel data is available, in order to have more robust results on market dynamics.

References

D. Acemoglu and P. Restrepo. The race between man and machine: Implications of technology for growth, factor shares, and employment. *American economic review*, 108(6):1488–1542, 2018.

T. Adamopoulos, L. Brandt, J. Leight, and D. Restuccia. Misallocation, selection, and productivity: A quantitative analysis with panel data from china. *Econometrica*, 90(3):1261–1282, 2022.

S. Ayerst, L. Brandt, and D. Restuccia. Market constraints, misallocation, and productivity in vietnam agriculture. *Food Policy*, 94:101840, 2020.

C. B. Barrett, S. M. Sherlund, and A. A. Adesina. Shadow wages, allocative inefficiency, and labor supply in smallholder agriculture. *Agricultural Economics*, 38(1):21–34, 2008.

D. Benjamin. Household composition, labor markets, and labor demand: testing for separation in agricultural household models. *Econometrica: Journal of the Econometric Society*, pages 287–322, 1992.

M. Bils, P. J. Klenow, and C. Ruane. Misallocation or mismeasurement? *Journal of Monetary Economics*, 124:S39–S56, 2021.

C. Chen, D. Restuccia, and R. Santaeulalia-Llopis. Land misallocation and productivity. Technical report, National Bureau of Economic Research, 2017.

B. Y. Coy Castellanos. Determinantes del acceso a crédito agropecuario en colombia. *Economía*, 2017.

J. G. Cragg. Some statistical models for limited dependent variables with application to the demand for durable goods. *Econometrica: Journal of the Econometric Society*, pages 829–844, 1971.

B. Dillon and C. B. Barrett. Agricultural factor markets in sub-saharan africa: An updated view with formal tests for market failure. *Food policy*, 67:64–77, 2017.

R. Eastwood, M. Lipton, and A. Newell. Farm size. *Handbook of agricultural economics*, 4:3323–3397, 2010.

A. D. Foster and M. R. Rosenzweig. Economic development and the decline of agricultural employment. *Handbook of development economics*, 4:3051–3083, 2007.

N. Garcés Rodríguez and H. Gomez. Cross-sectional productivity estimation: A practical guide. Technical report, Alianza EFI, 2022.

X. Giné. Access to capital in rural thailand: An estimated model of formal vs. informal credit. *Journal of Development Economics*, 96(1):16–29, 2011.

D. Gollin and C. Udry. Heterogeneity, measurement error, and misallocation: Evidence from african agriculture. *Journal of Political Economy*, 129(1):1–80, 2021.

D. Gollin, D. Lagakos, and M. E. Waugh. The agricultural productivity gap. *The Quarterly Journal of Economics*, 129(2):939–993, 2014.

J. Hamory, M. Kleemans, N. Y. Li, and E. Miguel. Reevaluating agricultural productivity gaps with longitudinal microdata. *Journal of the European Economic Association*, 19(3):1522–1555, 2021.

R. Martínez-Espiñeira. A box-cox double-hurdle model of wildlife valuation: The citizen's perspective. *Ecological Economics*, 58(1):192–208, 2006.

J. Merfeld. Misallocation and agricultural production: Evidence from india. 2021.

J. A. Ocampo. *Una historia del sistema financiero colombiano: 1951-2014*. Portafolio, 2015.

C. Oliveros, J. Sanz, C. Ramírez, and C. Tibaduiza. Ecomill: Tecnología de bajo impacto ambiental para el lavado del café. Technical report, Centro Nacional de Investigaciones de Café (Cenicafé), 2014.

D. Restuccia and R. Santaeulalia-Llopis. Land misallocation and productivity. *NBER working paper*, (w23128), 2017.

D. Restuccia, D. T. Yang, and X. Zhu. Agriculture and aggregate productivity: A quantitative cross-country analysis. *Journal of monetary economics*, 55(2):234–250, 2008.

C. Serna, J. Trejos, G. Cruz, et al. Estudio económico de sistemas de producción cafeteros certificados y no certificados en dos regiones de colombia. 2010.

F. Simtowe and M. Zeller. The impact of access to credit on the adoption of hybrid maize in malawi: An empirical test of an agricultural household model under credit market failure. 2006.

T. B. Wodjao. A double-hurdle model of computer and internet use in american households. *Departement of Economics, Western Michigan University. Fabrizio Carlevaro, Yves Croissant, Stéphane Hoareau*, 49, 2020.

8 Appendix

Table 8: Regression results for the OLS recursion test

	Harvest labor	Non-Harvest labor
Households with formal credit		
Household members	-0.13414	1.31313
	(0.31167)	(0.91054)
Prime female share	1.0157	-1.51475
	(1.21435)	(3.01664)
Prime male share	1.7061	2.86311
	(0.9323)	(2.92513)
Elderly female share	0.72514	0.22868
	(1.30067)	(3.31313)
Elderly male share	0.57272	0.75682
	(1.51874)	(3.06764)
Households with informal credit		
Household members	-0.91452	-0.53997
	(0.54843)	(1.20213)
Prime female share	3.29836	6.02303
	(1.75719)	(3.6926)
Prime male share	2.13739	5.58836
	(1.30995)	(3.40707)
Elderly female share	1.41454	2.06732
	(2.64037)	(5.00896)
Elderly male share	1.20859	6.70181*
	(1.330)	(3.13949)
Households with mixed credit		
Household members	0.79619*	0.71185
	(0.36782)	(1.1845)
Prime female share	-0.94919	2.17372
	(1.31782)	(4.35631)
Prime male share	-0.19276	3.33665
	(1.15856)	(3.79519)
Elderly female share	1.1072	4.11625
	(1.25259)	(5.5031)
Elderly male share	-1.14038	-1.93893
	(1.61101)	(4.52985)
Households without credit		
Household members	-0.25706	1.97498**
	(0.30054)	(0.75502)
Prime female share	1.77623	0.56012
	(1.18050)	(2.77226)
Prime male share	1.50262	-1.56373
	(1.11136)	(2.72515)
Elderly female share	0.91588	-0.86808
	(1.1315)	(2.65799)
Elderly male share	0.24244	0.71395
	(1.06408)	(2.62474)
Municipality FE	Yes	Yes
Control variables	Yes	Yes
R-squared	0.3689	0.225
N	525	525

NotesAll numbers in parentheses are clustered standard errors at the municipality level. All production factors are inverse hyperbolic sine transformed. Mixed credit status is households with both formal and informal credit. Significant variables in credit status (Elderly male share in informal credit and household members in mixed credit) is rejected when doing a F-test within credit status.

Table 9: Endogeneity test for the recursion test

	(1) Cultivated land	(2) Farm area	(3) Share of cultivated land over farm area
Household members	-0.12226	0.08448	-0.05863**
	(0.06571)	(0.10606)	(0.01991)
Prime male share	-0.44157	0.07902	-0.04184
	(0.29129)	(0.30469)	(0.05462)
Prime female share	-0.08013	0.54223	-0.00021
	(0.18994)	(0.28037)	(0.04683)
Elderly male share	-0.44999	0.50538	-0.10086
	(0.26467)	(0.34502)	(0.06110)
Farm area	0.87345***		
	(0.02566)		
Municipality FE	Yes	Yes	Yes
R-squared	0.6096	0.1955	0.132
N	525	525	525

Table 10: OLS results for machinery recursion test

	(1)
Cultivated land	0.16646**
	(0.05488)
Households with credit	
Household members	0.19693
	(0.28763)
Prime Male share	0.46033
	(0.50506)
Prime Female share	-0.37622
	(0.96103)
Elderly Male share	-0.08626
	(0.81563)
Elderly Female share	0.2288
	(0.60946)
Households without credit	
Household members	0.11987
	(0.10681)
Prime male share	0.55749
	(0.47384)
Prime female share	0.31826
	(0.48193)
Elderly Female share	0.28641
	(0.41339)
Elderly male share	0.22746
	(0.41368)
Municipality FE	Yes
R-squared	0.1157
N	525

Table 11: Variance Inflation factor for recursion test regression

	GVIF	Df	GVIF^(1/(2*Df))
ihs(Land)	1.341	1	1.158
ihs(Machines)	1.141	1	1.068
HHH sex	1.305	1	1.142
Distance to urban areas	1.768	1	1.330
Certified coffee	1.278	1	1.131
Average education	1.527	1	1.236
Farms Antiquity	1.458	1	1.208
MUNICIPIPALITY	5.586	28	1.031
credit:ihs(Total HH members)	10.675	1	3.267
credit:ihs(Prime males hare)	10.479	1	3.237
credit:ihs(Prime female share)	8.022	1	2.832
credit:ihs(Elderly female share)	3.458	1	1.860
credit:ihs(Elderly male share)	5.452	1	2.335
ihs(Total HH members):No credit	12.579	1	3.547
ihs(Prime males hare):No credit	10.128	1	3.182
ihs(Prime female share):No credit	8.632	1	2.938
ihs(Elderly female share):No credit	4.482	1	2.117
ihs(Elderly male share):No credit	6.041	1	2.458

Table 12: Recursion test for aggregated labor demand

	Dependent variable:
ihs(Land)	0.262^{**}
	(0.106)
ihs(Machines)	0.086
	(0.102)
Female HHH	−0.256
	(0.201)
distance to urban areas	-0.003^{*}
	(0.002)
Certified coffee	0.341^{*}
	(0.192)
Average education	0.071^{**}
	(0.029)
Farms Antiquity	−0.010
	(0.008)
credit:ihs(Total HH members)	0.164
	(0.240)
credit:ihs(Prime males hare)	0.904
	(0.797)
credit:ihs(Prime female share)	0.496
	(0.863)
credit:ihs(Elderly female share)	0.737
	(0.980)
credit:ihs(Elderly male share)	0.587
	(0.889)
ihs(Total HH members):No credit	0.384
	(0.286)
ihs(Prime males hare):No credit	−0.321
	(0.972)
ihs(Prime female share):No credit	0.819
	(1.033)
ihs(Elderly female share):No credit	0.054
	(0.982)
ihs(Elderly male share):No credit	−0.340
	(0.969)
Constant	3.260^{*}
	(1.918)
Note:	*p<0.1; **p<0.05; ***p<0.01