POLICY RESEARCH WORKING PAPER 4566

Reading Tealeaves on the Potential Impact of the Privatization of Tea Estates in Rwanda

B. Essama-Nssah
Kene Ezemenari
Vijdan Korman

The World Bank
Eastern Africa 2 Country Department
Poverty Reduction and Economic Management 3
March 2008

Policy Research Working Paper 4566

Abstract

The Poverty Reduction Strategy of the Government of Rwanda seeks to unlock the growth and poverty reduction potential of the tea sector through the privatization of tea estates. This paper uses the logic of causal inference and data from the 2004 Quantitative Baseline Survey of the tea sector to assess the potential impact of the privatization program. This entails a normalized comparison of productivity outcomes to account for household heterogeneity in terms of observable and non-observable determinants of these outcomes. The paper also compares living standards between tea and non-tea households. Three main findings emerge from the analysis. Productivity outcomes are generally better in the private sector than in the public sector. Male-headed households outperform female-headed households along all dimensions considered here. And tea households tend to be better off than non-tea households.

This paper—a product of the Poverty Reduction and Economic Management 3 Division, Eastern Africa 2 Country Department—is part of the series of analytical work feeding into the Poverty and Social Impact Analysis of Tea Sector Privatization in Rwanda that has also informed the Country Economic Memorandum, "Rwanda-Toward Sustained Growth and Competitiveness." Policy Research Working Papers are also posted on the Web at http://econ.worldbank.org. The authors may be contacted at Bessamanssah@worldbank.org, Kezemenari@worldbank.org, and Vkorman@worldbank.org.

Reading Tealeaves on the Potential Impact of the Privatization of Tea Estates in Rwanda

B. Essama-Nssah[•]
Kene Ezemenari
Vijdan Korman

The World Bank

[•] The authors are grateful to Kalpana Mehra for help with data exploration.

1. Introduction

The Poverty Reduction Strategy (PRS) of the Government of Rwanda, published in 2002, identifies rural development and agricultural transformation as the top priority (out of six) for promoting private sector-led development in that country[1]. This high focus on the rural economy is justified by the fact that agriculture contributes at least 40 percent of GDP and provides a livelihood for about 90 percent of the population. Therefore, growth in the agriculture is key to reducing the poverty rate of 60 percent of the population, based on a poverty line of 64, 000 RWF (about US $140) per person per year[2].

Agriculture also contributes significantly to Rwanda's trade with the rest of the world. In 2005, agricultural products accounted for just over 60percent of total exports in goods. Tea and coffee, the main cash crops, accounts for about 56 percent of these exports, and more than half of Rwanda's export revenue. Over the past 5 years, tea has remained, on average, the second major export crop (after coffee), although tea exports in some years (notably 2000 and 2001) have exceeded coffee exports. The sector is the largest employer in the country and directly generates close to 60,000 jobs.

Despite the growth in the sector following the genocide, several key problems limit the potential for this sector to generate foreign exchange and to contribute toward increased welfare of the population. The key problems constraining potential in the sector include: (i) agronomic conditions related to the location of factories and the quality and type of surrounding soil; (ii) low capacity of factories related to years of inadequate investment; (iii) differences in fertilizer application. With regard to the latter point, there have been reports that managers of the government owned factories apply less than optimum levels of fertilizer in order to ensure that production levels do not surpass the capacity of the tea processing factories. These problems have resulted in poor outcomes for the sector that are manifested in low producer prices, and low average yields. Based on these poor indicators, and the inefficiency of the government owned factories and plantations, the Government of Rwanda initiated a reform program based on

---

[1] The other five priorities include: human development, economic infrastructure, governance, private sector development and institutional capacity development.

[2] This is based on data from the 2001 Household Living Conditions Survey, also known as *Enquête Intégrale sur les Conditions de Vie* (EICV) *des Ménages au Rwanda*

privatization of the tea factories, to stimulate investment in the sector. Thus, in 1999, the government launched a phased privatization process that ensures a significant stake for tea growers and other local investors while attracting foreign investment as well.

The purpose of this paper is to provide a quantitative assessment of the likely impact of the privatization of tea estates in Rwanda, based on data from the 2004 Quantitative Baseline Survey of the Tea sector (QBST)[3]. The analysis is intended to serve as an input to the ongoing reform process.

The outline of the rest of the paper is as follows. Section 2 presents an overview of the tea sector and main issues. Section 3 presents our evaluation framework. Methods of impact evaluation are interpreted as ways of dealing with heterogeneity that may confound impact assessment. Such heterogeneity stems from observable and non-observable individual characteristics. Section 4 offers a discussion of the empirical results. It starts with a description of the underlying data. Then, it focuses on comparing outcomes among tea households. Finally, we compare living standards between these households and those not directly involved in the tea sector. Concluding remarks are made in section 5.

2. Overview of the Tea Sector

Tea is one of the two main export crops in Rwanda and has tremendous potential as a source of foreign exchange as well as a means of poverty reduction. It is grown on roughly 11,500 hectares of land on hills or drained marsh areas, which accounts for roughly 1 percent of the country's cropped area. Tea in Rwanda is mainly cultivated by small farmers, on a total surface area, per farm household that is less than 0.25 hectares. It is one of the few labor intensive crops that provide regular cash income to farmers, and employment opportunities to the general rural population. Until the onset of civil war and genocide of 1994, tea production had increased steadily.

[3] Also known as *Enquête Quantitative de Base auprès des ménages des zones Théicoles* (EQBT)

Tea production is organized around 11 estates distributed among 5 provinces[4] mostly in the western part of the country. An estate is a tea producing unit including a factory, a plantation (also known as *Bloc Industriel*), private tea plots and an associated forest to provide fuel wood to the factory for tea processing. Not all estates have all these components, for instance some own no plantations (World Bank 2003, p.34). The green leaves processed by a factory are supplied by the estate's plantation (if any) and independent tea growers working on individual plots with an average size of 0.25 ha. There are about 27,000 such independent growers owning nearly 70 percent of the total area under tea cultivation.

All growers belong to some organization either a *cooperative* when land is collectively owned, or an *association* based on private ownership of plots, or *thé villageois*, which refer to the thousands of small-holder producers engaged in green tea leaf production, and who do not form part of an association that supplies green leaf tea to the tea factories. There are only three cooperatives operating at Gisakura, Mulindi and Shagasha. Members of these cooperatives are paid a daily wage while growers who belong to an association earn an income directly from the parcel of tea they own (there are 13 growers' associations). In general, growers' organizations play a key management role in the process. They distribute fertilizer, collect and deliver tea leaves to the factory, pay the pluckers[5] and the growers themselves, and redistribute surplus earnings to members. It is estimated that a grower receives about 27 percent of the going price of a kilogram of leaves (12 out of 45 RWF). Besides pluckers, growers also employ unskilled workers or laborers for day-to-day maintenance tasks such as weeding and drainage. They are employed on a daily basis and earn on average 250 RWF per day (about 50 US cents). An umbrella organization FERWATHE (*Fédération Rwandaise des Théiculteurs* or Rwandese Federation of Tea Growers) was created in 2001 to protect the interests of growers in the new set of circumstances created by the liberalization process. All official organizations are members of this federation.

[4] (1) Byumba province: Mulindi, SORWATHE; (2) Cyangugu province: Gisakura, Nshili-Kivu, and Shagasha; (3) Gikongoro province: Kibati and Mata; (4) Gisenyi province: Nyabihu, Pfunda and Rubaya; (5) Kibuye province: Gisovu.

[5] Pluckers are skilled workers specialized in harvesting tealeaves. A pluck consists of the tea bud and one or two adjoining leaves and no more. Plucks should be delivered promptly to the factory for processing to avoid loss of quality through withering.

Three factories have been sold so far, aside from Government shares in SOWARTHE which were sold in 2003. SORWATHE[6], has always been under private control since its establishment in 1975. In February 2003, the Government sold its share of 23.54 percent, to the private company (13.54 percent) and to the association of tea growers (10 percent). A qualitative study conducted by the World Bank and the Government of Rwanda (World Bank 2003) to assess the likely poverty and social impacts of tea sector reforms noted that the yield of SORWATHE's plantations is about two and a half times higher than the average yield on state-owned estates (excluding Nshili-Kivu)[7]. Also, yields for the independent growers associated with the private estate, SORWATHE, are believed to be twice as high as the average from public estates. These observations provide a *working hypothesis* for our analysis, namely that *outcomes are expected to be better in the private sector than in the public sector*.

3. Accounting for Heterogeneity in Sectoral Outcome Comparison

To make meaningful comparisons of outcomes across sectors, we frame the analysis within the logic of *causal inference*. Indeed, the effect of a cause can be understood only in relation to another cause (Holland 1986). This idea is akin to that of assessing the return to a resource engaged in one activity relative to its opportunity cost, i.e. what the resource would have earned in the next best alternative use. In particular, for a tea household engaged in the private sector, we cannot assess the worth of the observed outcome without some information on the *counterfactual* i.e. what the household would have experienced had it been engaged instead in the public sector. Since we cannot observe a tea household engaged simultaneously in the private and public sectors, we construct the needed counterfactual from the information on the tea households engaged in the public sector. These counterfactual outcomes are constructed,

[6] SORWATHE stands for *Société Rwandaise du Thé*. The local name of the estate is Cyohoya-Rukeri. It was founded and is still owned by an American company, Tea Importers, Inc. of Westport Connecticut. Its plantations cover about 2 percent (or 252 ha) of the total area under tea cultivation. It is reported that this estate and the associated growers apply substantially more fertilizer than other estates.

[7] State ownership is managed by the Tea Board known as OCIRTHE an off-shoot OCIR (*Office des Cultures Industrielles du Rwanda*) which used to cover both tea and coffee.

using standard methods of non-experimental impact analysis, in a way that allows us to attribute the net outcome to participation in the private sector.

The methodological issue we face here is to find a way of assessing the *payoff* from participation in a social arrangement. For instance, if we observe that yields are higher for tea growers in the private than in the public sector, to conclude that participation in the private sector is better than in the public sector our method of comparison must control for any other factor (besides participation in the private sector) that can influence the outcome of interest. The logic of causal inference requires a model that explains both the process that sorts individuals between the two states of nature (participation versus nonparticipation) and the conditional outcomes. This section reviews the standard non-experimental methods that we use in this study, namely *matching* methods and *regression* analysis. We start the discussion with a benchmark case where agents are assumed homogenous with respect to all other dimensions besides participation.

The Benchmark Case of Unit Homogeneity

In general, the unit of analysis could be an individual, a household, a village, or a broader community such as a district or a province. Let the variable **y** stand for the outcome of interest (e.g. yield, cost of production or expenditure per capita). The effect of participation (akin to that of exposure to an intervention) on unit **i**, (call it $\mathbf{g_i}$) is measured relative to nonparticipation (non-exposure) on the basis of the outcome variable. Formally, we write $g_i = (y_{1i} - y_{0i})$, where $\mathbf{y_{1i}}$ is the observed outcome under participation and $\mathbf{y_{0i}}$ is the counterfactual. It is impossible to observe the value of the response variable for the same individual under two mutually exclusive states of nature (exposure and non-exposure). This is why evaluation methods are considered as ways of dealing with this *missing data* problem. If the intervention is limited to a subset of the population as is the case here, many of the methods suggest turning to non-exposed units (non-participants) in search of the missing information. They also specify circumstances under which the use of such information yields reliable estimates of the relevant effect.

The assumption of *unit homogeneity* (Holland 1986) characterizes a benchmark case where the effect on individual **i** could be reliably estimated. An individual response is a function of participation, observable and unobservable characteristics. Suppose we can find among non-participants an individual **j** with the same pre-exposure (observable and non-observable) attributes as participant **i**. Thus, under unit homogeneity, the outcome of this non-participant is a proxy for what would have happened to **i** had she not received the intervention. Hence, the effect of the intervention on **i** can be estimated as: $g_i = (y_{1i} - y_{0j})$.

The assumption of unit homogeneity is thus analogous to the *ceteris paribus* assumption used in scientific enquiry. The assumption serves as a benchmark case against which to assess the implications of *heterogeneity*. In non-exposure state, one would generally expect response heterogeneity for participants and non-participants, particularly when eligible candidates are given the choice to participate or not[8]. Such heterogeneity can confound impact assessment, leading to biased results. We now review briefly matching and regression methods of controlling for heterogeneity.

Matching Methods

If the mechanism that sorts individuals among sectors (i.e. states of nature) is based exclusively on observable characteristics[9], then the counterfactual outcome for participant **i** would be equal to the outcome of nonparticipant **j** with the same observables. Exact matches are usually difficult to find, thus we may tolerate some deviation from sameness and consider nonparticipants who are almost like the participant under consideration (a sort of second best solution). Let **z** stand for the set of observable characteristics of participant **i**. We can think of a tolerance criterion as a cut-off distance

[8] Heckman and Smith (1995) cite the case where those who choose to join a social program do so because of the poor alternative they face outside the program. In such a case, non-participants would have better outcomes than participants had the latter not elected to participate. This response heterogeneity is also known as *selection bias*.

[9] This case is known as the assumption of *conditional independence*. After conditioning on observable characteristics, the absence of unobservable heterogeneity between participants and nonparticipant implies that any systematic differences in outcomes between the two groups are due to participation. One rendition of the same assumption states that: given observable characteristics, potential outcomes are independent of participation.

defining a neighborhood of **z** in the space of attributes such that any nonparticipant **j** with a set of attributes in that neighborhood qualifies as a look-alike for **i**.

In practice matching may become more and more difficult, the larger the set of observable characteristics underpinning the matching exercise. Rosenbaum and Rubin (1983) show that the *dimensionality* of the problem can be significantly reduced by matching on the propensity score[10]. Thus instead of conditioning on an n-dimensional variable, units are matched on a scalar variable. This simplification is possible because conditional independence remains valid if we use the propensity score **p(z)** instead of the covariates **z**.

The computation of the counterfactual outcome for any participant **i** with propensity score $\mathbf{p_i}$ entails three basic steps: (1) Use a measure of *proximity* to identify nonparticipants in the comparison group whose scores are close enough to $\mathbf{p_i}$ [all observations satisfying this condition belong to a neighborhood $\mathbf{c(p_i)}$]; (2) Select a weighing function that assigns some weight to each member of $\mathbf{c(p_i)}$ in the computation of the counterfactual outcome for participant **i**; (3) Compute the counterfactual outcome as a weighted average of the outcomes of members of $\mathbf{c(p_i)}$ according to the following expression.

$$\hat{y}_i = \sum_{j \in c(p_i)} w_{ij} y_j ; \; w_{ij} \in [0, 1]; \quad \sum_{j \in c(p_i)} w_{ij} = 1 \tag{3.1}$$

The feasibility of this approach requires an overlap between the distribution scores of participants and that of nonparticipants. The fuller the overlap, the easier it is to find matches. This is why, in practice, matching is usually restricted to the *region of common support*.

Expression (3.1) reveals that the counterfactual outcome for participant **i** is computed as a *locally weighted average* or a *moving average* of relevant outcomes in the comparison group. One can think of this procedure as sliding a window of a given width across the space of scores of nonparticipants and taking the average of the outcome variable for all observations in the window. Furthermore, it is well known that the mean

[10] This result is the foundation of the popular method of impact evaluation known as propensity score matching (PSM). The propensity score is the conditional probability of participation given the observed attributes.

of a variable can also be computed by running a regression of the variable on a constant. In other terms, the locally weighted average estimator of the counterfactual outcome for participants **i** is also a *locally weighted regression*. A semi-parametric extension of this idea is based on the following considerations.

Assume that the outcome of nonparticipant **j** is a separable function of observables as summarized by the propensity score $\mathbf{p_j}$, and unobservable characteristics represented by the random disturbance, $\mathbf{u_j}$. Thus we write: $y_j = \beta(p_j) + u_j$. If the expected value of the random disturbance is zero, then Taylor's expansion allows us to write the expected outcome near $\mathbf{p_i}$ (the score of participant i) as follows.

$$\beta(p_j) \approx \beta_0 + (p_j - p_i)\beta_1 \tag{3.2}$$

Locally weighted regression minimizes the following weighted sum of squares.

$$S(\beta) = \sum_{j=1}^{n} w_{ij}\left[y_j - \beta_0 - (p_j - p_i)\beta_1\right]^2 \tag{3.3}$$

Hence, the outcome participant **i** would have achieved had she not participated in the arrangement is equal to:

$$\hat{y}(p_i) = \hat{\beta}_0 \tag{3.4}$$

Note that the estimate varies with location (i.e. $\mathbf{p_i}$). This process must be repeated for each participant[11].

As far as the choice of weights is concerned, one can follow the nearest-neighbor approach or use a kernel function. For each participant **i**, the nearest-neighbor method searches for the nonparticipant **j** with the closest propensity score to **i**. This nonparticipant gets a weight of 1 and all others get a weight of zero. When there are many candidates, the method assigns equal weight to each and zero to nonparticipants

[11] Smith and Todd (2005) explain that matching by local linear regression is helpful in situations where the distribution of observations from the comparison group around a given participant is asymmetrical as in the case where there are gaps in the distribution of propensity scores.

outside the neighborhood **c(p$_i$)**. The weights associated with a kernel function are defined as follows.

$$w_{ij} = \frac{K\left(\frac{p_i - p_j}{h}\right)}{\sum_{j \in \{d=0\}} K\left(\frac{p_i - p_j}{h}\right)} \tag{3.5}$$

where **h** stands for the tolerance level (also known as bandwidth), and the set **{d=0}** represents the comparison group. Our analysis is based on the Gaussian kernel[12].

Individual gains from participation can now be written as:

$$g_i = (y_i - \hat{y}_i) = \left(y_i - \sum_{j \in c(p_i)} w_{ij} y_j \right) \tag{3.6}$$

These are the basic ingredients for the computation of an impact indicator. The most commonly used indicator is the mean gain from participation[13]. It is equal to:

$$\theta_M = \sum_{i \in T} \omega_i \left(y_i - \sum_{j \in c(p_i)} w_{ij} y_j \right) = \sum_{i \in T} \omega_i g_i \tag{3.7}$$

Where **T** stands for the set of participants (i.e. the treated), and $\boldsymbol{\omega_i}$ can be interpreted more broadly as the evaluative weight assigned to participant **i**. In standard applications, $\boldsymbol{\omega_i}$ is taken to be the sampling weight associated with observation **i**. To look beyond this average impact one can plot $\mathbf{g_i}$ or the ratio of the observed outcome (y_i) to the counterfactual ($\hat{y}_i$) as a function of **q**, the cumulative distribution of the participants ranked in increasing order of some variable (e.g. the counterfactual outcome). Participation would have a positive impact at each percentile where $\mathbf{g_i}$ is greater than zero or the ratio is greater than one. Such plots are known as *Program Incidence Curves*[14].

[12] Other possible choices include: Epanechnikov, bi-weight or quartic, triangular, tri-weight, uniform, and cosinus.

[13] This indicator is also known in the literature as the average treatment effect on the treated (ATET).

[14] More generally, we may also refer to these as *Participation Incidence Curves*. They reveal the differential gains (or losses) from the participation in a social arrangement.

Regression Analysis

Regression analysis can also be used to control for heterogeneity. Let $y_{1i} = \beta_1(x_i) + u_{1i}$ be the outcome if unit i participates in the arrangement, and $y_{0i} = \beta_0(x_i) + u_{0i}$ the outcome in the nonparticipation state. Let $\mathbf{d_i}$ be an indicator of participation which is equal to 1 in the participation state and 0 otherwise. The potential outcome for any unit can therefore be written as: $y_i = d_i y_{1i} + (1 - d_i) y_{0i}$. This is equivalent to the following general expression.

$$y_i = \beta_0(x_i) + [\beta_1(x_i) - \beta_0(x_i) + (u_{1i} - u_{0i})]d_i + u_{0i} \tag{3.8}$$

Smith and Todd (2005) interpret the above equation as a random coefficient model, because the effect of participation varies across individuals even if we control for observable characteristics $\mathbf{x_i}$. We get the fixed coefficient or common effect version of the model if we make the following two assumptions: (1) Unobservable characteristics are the same in the participation and nonparticipation states; (2) The function $\theta(x_i) = [\beta_1(x_i) - \beta_0(x_i)]$ is constant with respect to observable characteristics. If, in addition we assume that $\boldsymbol{\beta}(\mathbf{x_i})$ is linear in parameters, then we get the familiar expression of the common effect model.

$$y_i = x_i\beta + \theta d_i + u_i \tag{3.9}$$

If conditional independence prevails, then $\mathbf{d_i}$ and $\mathbf{u_i}$ are independent given $\mathbf{x_i}$. OLS provides a consistent estimate of average impact, $\hat{\theta}$. This is the parametric equivalent of matching estimates.

If conditional independence fails so that $\mathbf{d_i}$ is correlated with $\mathbf{u_i}$, then Heckman's selection estimate of average impact can be obtained by applying OLS to the following equation (LaLonde 1986):

$$y_i = x_i\beta + \theta d_i + \sigma_{u\varepsilon}\hat{\lambda}_i + \nu_i \tag{3.10}$$

where $\hat{\lambda}_i = [d_i \hat{\lambda}_{1i} + (1-d_i)\hat{\lambda}_{0i}]$ is an estimate of the inverse Mills ratio derived from a probit model of participation. The coefficient of this variable is a function of the covariance between unobservables in the participation model (**ε**) and those in the outcome equation (**u**).

To relax the assumption that $\theta(x_i) = \theta$, we can apply the Heckman's procedure separately to participants and nonparticipants. In the first case, the estimating equation is:

$$y_{1i} = x_i\beta_1 + \sigma_{1\varepsilon}\hat{\lambda}_{1i} + \nu_{1i}, \forall d_i = 1 \tag{3.11}$$

and for nonparticipants:

$$y_{0i} = x_i\beta_0 - \sigma_{0\varepsilon}\hat{\lambda}_{0i} + \nu_{0i}, \forall d_i = 0 \tag{3.12}$$

Estimating separate outcome equations also allows us to compute individual gains from participation as follows (Maddala 1983).

$$g_i = x_i\left(\hat{\beta}_1 - \hat{\beta}_0\right) + (\hat{\sigma}_{1\varepsilon} - \hat{\sigma}_{0\varepsilon})\hat{\lambda}_{1i} \tag{3.13}$$

The Heckman approach is a two-stage procedure that treats unobservable heterogeneity as a *problem of an omitted variable*. The proposed solution is to include an estimate of the omitted variable as an explanatory variable in the outcome equation[15].

4. Estimates of Potential Impacts

In this section we estimate the potential impact of the privatization of tea estates in Rwanda. The outcomes of interest are determined on the basis of policy concerns. A fundamental expectation of the stakeholders is the privatization process will eventually lead to improved productivity and living standards for those engaged in the sector. We proceed in three steps. First we give a brief description of the sample we use in estimation. Then we focus our attention to productivity issues by considering only tea-

[15] One can also resort to the instrumental variable (IV) approach to deal with unobserved heterogeneity. This method relies on an exclusion restriction that assumes that there is at least one variable that determines participation but does not affect outcomes. This instrument can then substitute for $\mathbf{d_i}$ in equation (2.9) to restore some sort of conditional independence. Subsequent application of OLS would produce a consistent estimate of average impact. In general, one can turn to geography, politics or discontinuities created by program design in search of suitable instrumental variables (Ravallion 2005).

growing households, and comparing yield and cost elements between the private and public sector. Finally, we use the full sample to compare economic welfare between households engaged in the sector and those who are not.

Data

Our empirical analysis is based on the QBST, a baseline survey conducted in 2004. It is part of a planned series of surveys designed to monitor the productivity and the living standard of the populations engaged in the tea sector. It is important to keep in mind the survey was taken before the implementation of the privatization. That is why we speak of potential impact. The survey provides information on three basic dimensions of interest: (1) productivity indicators such as yield, use and cost of fertilizer; (2) living standard as indicated by income and expenditure; and (3) access and use of social services.

The available sample includes about 2, 064 households representing the 102,812 households living in parts of the country where tea is grown. Thus each household in the sample stands for about 50 households for a total population of 515,217 inhabitants in tea-producing provinces. The average household size is about 5 people. It is estimated that only 30 percent of members of tea households are engaged in the tea sector. The rest is employed in non tea activities.

Table 1 Average Characteristics of Tea Growers

Characteristics	Private[16]	Public	All
Age	51.93	47.69	48.20
Female	0.22	0.30	0.29
Land	18.40	30.80	29.30
Livestock	0.37	0.48	0.47
Bicycle	0.16	0.06	0.07
Water30	0.42	0.28	0.30
Market30	0.20	0.15	0.15
Road30	0.40	0.22	0.24
Per Capita Expenditure	1640.39	2377.36	2288.20
Sample Size	83	603	686

Source: Authors' calculations

[16] These households are SORWATHE supported tea growers.

Figure 1a: Distribution of Propensity Scores Private Sector Growers

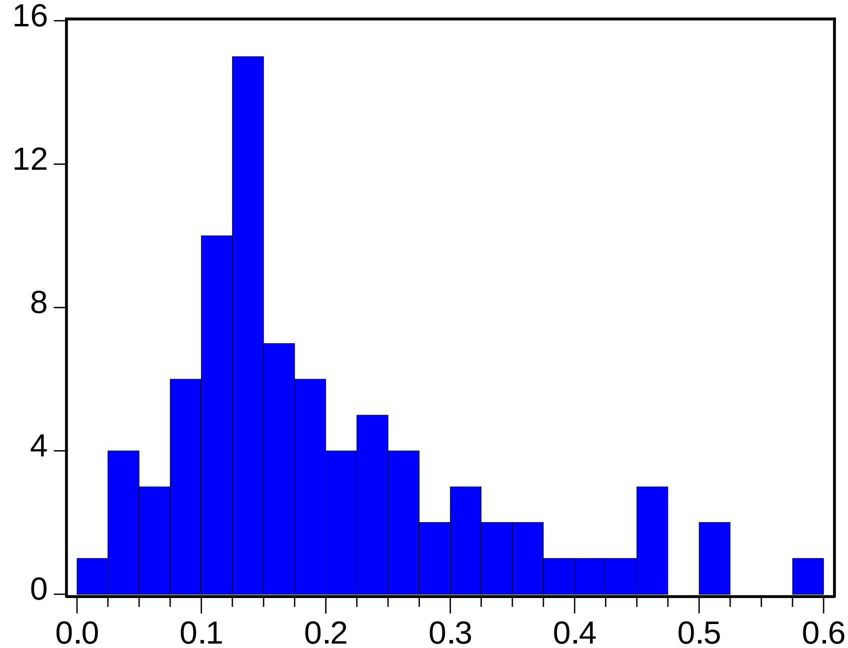
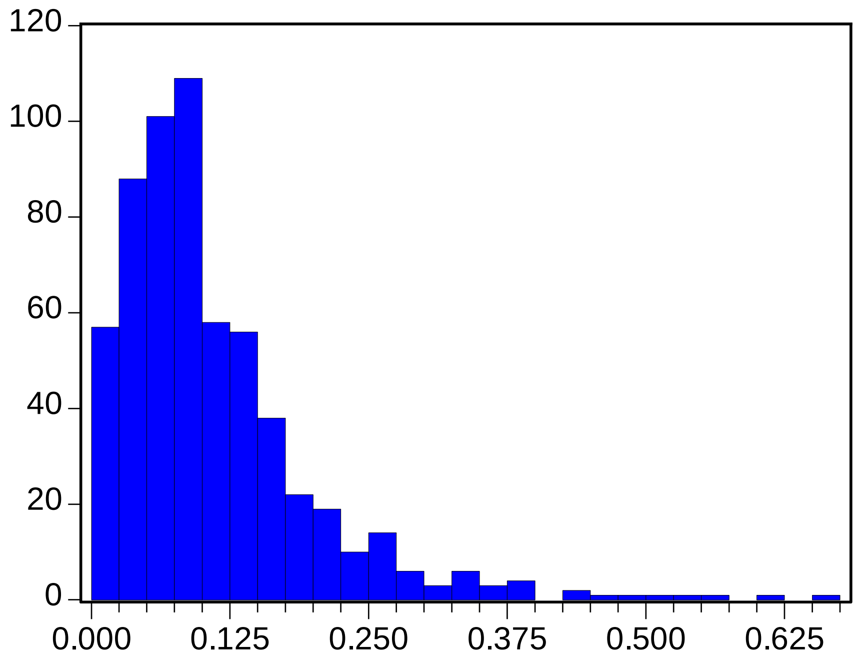


Figure 1b: Distribution of Propensity Scores for Public Sector Growers



Our analysis is based on the comparison of outcomes for the tea households supported by SORWATHE with those obtained by households dealing with the public sector. The quality of the conclusions stemming from this comparison hinges on the extent to which both groups are homogenous. Table 1 shows average values for nine observable characteristics of tea households. It is evident that these two groups differ significantly along those dimensions. For instance, the average age for SORWATHE households is 52 versus 47 years for households of the public sector. Average land holding is higher in the public sector (31 Ares[17]) than in the private sector (about 18). The data also reveal that about 16 percent of the households in the private sector own a bicycle versus 6 percent in the public sector. About 42 percent of the households in the private sector are less than 30 minutes from a water source compared to 28 percent in the public sector, similarly for the distance from a road. Per capita expenditure for the comparison group is about 45 percent higher than the average expenditure for the private sector households.

Table 2. Estimates from the Logit Model of the Propensity Score

Variable	Coefficient	Std. Error	z-Statistic	Prob.
Constant	-2.547092	0.498220	-5.112381	0.0000
Age	0.023063	0.008074	2.856644	0.0043
Female	-0.484354	0.301886	-1.604430	0.1086
Land	-0.009503	0.005706	-1.665551	0.0958
Livestock	-0.080630	0.154890	-0.520564	0.6027
Bicycle	1.422839	0.398861	3.567253	0.0004
Water30	0.523196	0.255485	2.047855	0.0406
Market30	0.387526	0.317964	1.218774	0.2229
Road30	0.618112	0.261701	2.361905	0.0182
Per Capita Expenditure	-0.000418	0.000125	-3.342370	0.0008

Source: Authors' calculations

The above noted heterogeneity constraints our ability to find in the comparison group, households similar to those associated with SORWATHE. As it can be seen in table 2, the characteristics for which the two groups differ the most tend to have significant coefficients in the logit aggregation function we use to match households on observables. The extent of this heterogeneity is also reveals the histograms of propensity

[17] One unit 'Are' is 100 squared-meter and one hectare (ha) is 10,000 squared-meter. Therefore, 100 Are is equal to 1 hectare.

scores presented in figure 1. It can be seen that the two histograms overlap most at lower levels of the propensity scores. Given this situation, we impose a much tighter level of tolerance in matching. For kernel matching we set the bandwidth at 0.01.

The Returns to Participation in the Private Sector

Given the current organization of tea production in Rwanda, do tea households operating within the SORWATHE system have better outcomes than the rest? To answer this question, we consider outcome differentials between the private and public sector along five dimensions. The first two, yield per hectare and time taken to carry leaves to the collection point are indicators of productivity. The yield is measured in kilograms (KG) of green tea per hectare while the time is measured in minutes. The other tree dimensions are related to the cost of production. They measure the use of fertilizer in KG per hectare, the cost of fertilizer in RWF per KG, and the cost of extension services per Are. Table 3 shows a comparison of mean outcomes between the private and public sector. This comparison does account for the heterogeneity among households.

Table 3. A Naïve Comparison of Outcomes across Sectors

Outcomes	Difference in Means	Private	Public	ALL
Yield per Hectare	869.32	9315.10	8445.78	8555.53
Time to Carry Leaves	-13.03	15.22	28.25	26.63
Fertilizer Use	-176.64	513.67	690.31	664.17
Fertilizer cost	-45.00	180.35	225.34	218.71
Extension Cost per Are	577.67	876.16	298.51	371.32
Sample size	-	83	603	686

Source: Authors' calculations.

The above results suggest that outcomes in the private sector are potentially better in the private sector than in the public sector. On average, private sector households have higher yield than public sector ones. They also take less time to carry leaves to the collection point, use less fertilizer and pay less for it than the comparison group. Private sector households pay more for extension services than public sector ones. To what extent do these conclusions stand up to *normalization on observables*?

Table 4 presents normalized impact estimates based on propensity score matching. Given that estimates depend crucially on the choice of weights, we compute the estimates using five different kernel functions[18]. All five kernel functions lead to results that are very close to each other. Except for the cost of fertilizer, the normalized estimates confirm the qualitative conclusions derived from the naïve outcome comparison. The normalized comparison reveals the tea growers associated with the private sector do pay slightly more for fertilizer than those in the public sector. They certainly have much higher yield per ha than the public sector households. In fact the normalized impact estimated for yield is roughly twice the naïve one. Yet, the better performing tea growers also use less fertilizer than the comparison group. This result suggests the possibility of inefficiencies in public extension services. It also suggests that the little extra cost for those services that the private sector households are paying per are may be worth it. Finally, it is likely that private sector growers produce better quality leaves given that they take about 12 minutes less than the comparators to carry leaves to the collection point.

Table 4. Accounting for Observable Heterogeneity in Outcome Comparison

	Yield	Fertilizer Use	Fertilizer Cost	Time to Carry	Extension Cost
Gauss	1768.16	-19.34	34.89	-11.58	576.32
Epanechnikov	1746.31	-58.59	32.55	-11.84	565.54
Quartic	1667.16	-68.94	31.52	-12.01	571.66
Uniform	1823.47	-47.16	32.73	-11.69	550.49
Cosinus	1732.39	-60.52	32.39	-11.87	566.70

Source: Authors' calculations

Table 5. Gender Differences in Yield

	Gauss	Epanechnikov	Quartic	Uniform	Cosinus
Female	1408.92	1172.29	1016.11	1329.53	1142.99
Male	1879.64	1927.58	1872.75	1979.45	1918.51
All	1768.16	1746.31	1667.16	1823.47	1732.39

Source: Authors' calculations

[18] The quartic kernel function is also known as the bi-weight kernel. Also note that the use of the uniform kernel is equivalent to radius matching, a variant of the nearest-neighbor method.

We now consider the gender dimension of some of these results. In general, male-headed households outperform female-headed households along all dimensions considered here, (note that the comparison is between females or males in the private sector and their nearest neighbors, regardless of whether male or female). We report only the most striking differences. Table 5 and table 6 show differences in yields and with respect to the use of fertilizer between male and female-headed households. It appears that men have yields that are much higher than women's. Yet, the former also use significantly less fertilizer than the latter. Could there be a gender bias in the private sector's extension services?

Table 6. Gender Differences in the Use of Fertilizer

	Gauss	Epanechnikov	Quartic	Uniform	Cosinus
Female	84.74	112.11	103.15	106.77	110.41
Male	-52.80	-114.45	-125.27	-97.54	-116.46
All	-19.34	-58.59	-68.94	-47.16	-60.52

Source: Authors' calculations

Looking Beyond the Tea Sector

Up to now, we have focused our attention on tea households, comparing outcomes for those engaged in the private sector with outcomes observed in the public sector. The development of the tea sector is a key element of the agricultural policy in support of the poverty reduction strategy in Rwanda. A recent analysis of the 2001 household survey by Dabalen et al. (2004) reveals that agriculture remains the principal source of earnings for the poor and that non-poor households are more likely than poor households to have earnings from non-farm activities. In this perspective, we analyze the available data to determine whether, other things being equal, tea households are better off than households who earn their living mostly from non-tea activities. We proceed in a manner that is entirely analogous to the way we compared outcomes within the tea sector. We use a set of observable characteristics to attenuate some of the bias due to such characteristics.

Table 7. Average Characteristics for Tea and Non-Tea Households

Characteristics	Tea	Non-Tea	ALL
Age	44.99	46.30	45.68
Education (years)	2.89	2.43	2.65
Male	0.75	0.67	0.71
Household Size	5.19	4.84	5.01
Land (ares)	20.37	0.02	9.83
Livestock	3.49	3.01	3.25
Bicycle	0.07	0.04	0.05
Road30	0.24	0.22	0.23
Water30	0.28	0.24	0.26
Sample Size	986	1053	2039

Source: Authors' calculations

Table 7 presents some average characteristics for 986 tea households and 1053 non-tea households. The most striking difference between these two groups relates to land ownership. Average landholding among tea households is more than a thousand times the average for non tea households. As one would expect, land ownership is a key determinant of participation in the tea sector. This fact is confirmed by the estimation results of a logit model of participation presented in table 8. In this model, land ownership has a very high level of statistical significance. Beyond land ownership, these results also indicate that gender (i.e being male) and years of education have a significant and positive impact on the likelihood that a household is engaged in the tea sector.

Table 8. A Model of Participation in the Tea Sector

Variable	Coefficient	Std. Error	z-Statistic	Prob.
Constant	0.277728	0.531392	0.522642	0.6012
Age	-0.063520	0.025603	-2.480920	0.0131
Age Squared	0.000333	0.000265	1.253871	0.2099
Years of Education	0.121048	0.049172	2.461716	0.0138
Years of Education Squared	-0.016652	0.005332	-3.122850	0.0018
Male	0.513665	0.160782	3.194791	0.0014
Household Size	0.069233	0.039357	1.759115	0.0786
Land Area	1.413235	0.178340	7.924366	0.0000
Livestock	-0.029024	0.019176	-1.513571	0.1301
Bicycle	0.472874	0.294378	1.606347	0.1082

Source: Authors' calculations

To what extent, if at all, are tea households better off than non-tea households? We base our answer to this question on several types of comparisons. Table 9 presents results from a naïve welfare comparison based on both per capita expenditure and per capita income. As noted earlier this type of comparison does not account for any heterogeneity between the two groups. The results suggest that, in tea cultivating regions of Rwanda, average welfare is higher for tea households than for non-tea households.

Table 9. Naïve Comparison of Welfare between Tea and Non-Tea Households

Outcome	Difference in Means	Tea	Non-Tea	All
Per capita Expenditure	1936.43	25397.66	23461.23	24399.25
Per capita Income	2571.38	23075.70	20504.32	21721.56

Source: Authors' calculations

Table 10. Matching Comparison of Welfare between Tea and Non-Tea Households

Kernel	Per Capita Expenditure	Per Capita Income
Gauss	6504.71	5488.51
Epanechnikov	6765.65	6659.29
Quartic	6722.41	6628.20
Uniform	6831.99	6770.90
Cosinus	6757.93	6651.70

Source: Authors' calculations

Next we use the regression methods described above in order to account for both observable and non-observable heterogeneity. In the case of the Heckman method, the selectivity correction factor turned out not to be statistically significant. This gave us comfort in our use of the propensity score matching method. The corresponding results are presented in table 10. These reveal that when likes are compared with likes, the welfare advantage that tea households have over the non-tea households is much higher than what the naïve comparison would suggest. Indeed, regardless of the kernel function used among the ones reported in table 10, average per capita expenditure for tea households is more than the average for non-tea households by about 7,000 RWF

As a last test for the robustness of our conclusion, we use a two-stage procedure explained by Wooldridge (2002). First estimate the participation equation as a nonlinear binary response model using the probit or logit model, just as in the first stage of propensity score matching Then use the estimated propensity score as an instrument for the participation indicator in the outcome equation and run OLS to estimate average impact. The results of this procedure applied to per capita expenditure are presented in table 11. They show that average difference in welfare between the two groups is statistically significant and equal about 3,643 RWF in favor of tea households.

Table 11. Regression Estimation of Average Difference in Per Capita Expenditure (Instrumental Variable Method)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
Constant	26518.49	2816.683	9.414794	0.0000
Male	2425.350	1168.025	2.076455	0.0380
Age	72.04438	33.89080	2.125780	0.0336
Education	1362.314	169.3803	8.042931	0.0000
Household size	-1570.103	262.6600	-5.977702	0.0000
Livestock	2219.420	293.1086	7.572006	0.0000
Market30	-6361.600	2183.026	-2.914120	0.0036
Market60	-3686.614	1985.779	-1.856508	0.0635
Market90	-5254.870	1981.555	-2.651892	0.0081
Market90P	-5949.287	2009.430	-2.960683	0.0031
Road30	-2144.925	1193.769	-1.796767	0.0725
Water30	-974.0571	1140.822	-0.853821	0.3933
Propensity Score	3643.693	1452.421	2.508703	0.0122

Source: Authors' calculations

4. Concluding Remarks

As one of the two main export crops in Rwanda, tea is a significant source of foreign exchange and potentially an important means of poverty reduction. It is in fact one of the few labor intensive crops that provide regular cash income to farmers and employment opportunities to some of the rural population. The Poverty Reduction Strategy of the Government of Rwanda seeks to unlock this potential by reforming its agricultural policy in general while focusing particularly on the key factors that constraint growth in the tea sector. An important component of this program of reforms involves the privatization of tea factories.

This paper uses data from the 2004 Quantitative Baseline Survey of the Tea sector to assess the potential impact of privatization of tea estates. The analysis is framed within the logic of causal inference. This entails a normalized comparison of outcomes to account for household heterogeneity in terms of observable and non-observable determinants of the outcomes of interest. These outcomes relate to productivity. Three main findings emerge from this comparison. Productivity outcomes such as yield, time taken to carry leaves to the collection point, and fertilizer use are generally better in the private sector than in the public sector. Also, male-headed households out perform female-headed households along all dimensions consider here. Finally, in a welfare comparison between the tea and non-tea sectors, the former tend to be better off than the latter.

References

Brambilla, Irene and Porto, Guido G. 2005. Farm Productivity and Market Structure: Evidence from Cotton Reforms in Zambia. Discussion Paper No.919. New Haven: Economic Growth Center (Yale University).

Dabalen, Andrew, Paternostro, Stefano and Gaëlle, Pierre. 2004. The Returns to Participation in the Nonfarm Sector in Rural Rwanda. World Bank Policy Research Working Paper No. 3462.

Diop, Ndiame, Brenton, Paul and Asarkaya, Yakup. 2005. Trade Costs, Export Development and Poverty in Rwanda. World Bank Policy Research Paper No. 3784.

Essama-Nssah, B. 2006. Propensity Score Matching and Policy Impact Analysis: A Demonstration in EViews. World Bank Policy Research Working Paper No. 3877.

Holland, Paul W. 1986. Statistics and Causal Inference. Journal of the American Statistical Association, Vol. 81, No. 396: 945-960.

LaLonde, Robert J. 1986. Evaluating the Econometric Evaluations of Training Programs with Experimental Data. The American Economic Review, Vol. 76, No. 4:604-620.

Loader, Catherine. 2004. Smoothing: Local Regression Techniques. In James Gentle, Wolgang Hardle, Yoichi Mori (eds): *Handbook of Computational Statistics*. Heidelberg: Springer-Verlag.

Maddala, G. S. 1983. Limited-Dependent and Qualitative Variables in Econometrics. Cambridge: Cambridge University Press.

Ravallion, Martin. 2005. Evaluating Anti-Poverty Programs. World Bank Policy Research Paper No. 3625. Washington D.C: The World Bank.

Rosenbaum, Paul and Rubin, Donald. 1983. The Central Role of the Propensity Score in Observational Studies for Causal Effects. Biometrika, Vol.70, No.1:41-55.

Smith, Jeffrey A. and Todd, Petra E. 2005. Does Matching Overcome LaLonde's Critique of Nonexperimental Estimators? Journal of Econometrics, Volume 125, No. 1-2: 305-353.

Wooldridge, Jeffrey M. 2002. Econometric Analysis of Cross Section and Panel Data. Cambridge (Massachusetts): The MIT Press.

World Bank. 2003. Rwanda Tea Sector Reform: Analysis of Poverty and Social Impacts — Stakeholder and Institutional Analysis. Washington, DC. The World Bank (mimeo)