

farmer by less than two years. The effects are relatively similar when calculated on the all sample. The time ratios of the two distances are statistically different at 10%.

Another interesting result deals with the effect of the level of education: four more years of formal education (e.g. from primary to secondary school) decreases the time length before adoption by slightly more than one and a half year for the adopters and by more than two years when the effect is calculated on the all sample. These results can be compared to those on learning externalities: farmers benefiting from more adopters in the neighbourhood (10%) need, on average, two years less to adopt.

**Table 7: Estimated effect of a variation in the covariate on the time length before adoption (the table indicates the difference in years)**

Variables	Variation in the covariate	Estimated mean effect on adopters <sup>1</sup>	Estimated median effect on all farmers <sup>2</sup>
<i>Household human capital assets</i>			
years of formal education	4 years	-1.7	-2.1
dependency ratio <sup>3</sup>	0.1	5.1	6.3
<i>Household physical assets</i>			
total acreage owned	1 acre	-0.3	-0.4
predicted probability for access to credit	10%	-0.8	-1.0
total rainfall since farm establishment	10% <sup>4</sup>	-4.7	-5.8
<i>External factors</i>			
distance to Nairobi on dry weather roads	10 km	10.8	12.8
distance to two other urban centres on all weather roads	10 km	1.7	2.1
% of "adopters" in the division	10%	-1.8	-2.2
annual precipitation / potential evapotranspiration ratio	10% <sup>4</sup>	-1.3	-1.6

<sup>1</sup> Mean duration for adopters = 5.7

<sup>2</sup> Median duration for all the farmers (adopters and non- adopters) = 7

<sup>3</sup> Number of children/ family size

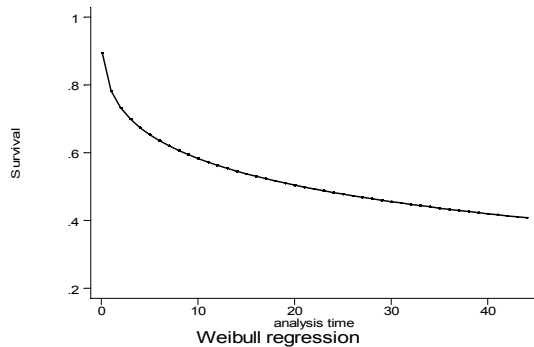
<sup>4</sup> 10% of the maximum amplitude of the variable.

For dummy variables, the effect of significant variables is presented using graphical representations of the estimated survival functions (using the results of the weibull specification of table 6). Graphs 3 below present the survival distributions before and after 1990, year considered as the start of the difficulties for dairy co-operatives and the liberalisation process. The graph shows that the survival "as non- adopters" is higher after 1990: farmers thus take more time to adopt after 1990, compared to before. While this result suggests that the adoption trend has slowed down after the liberalisation, caution is to be exercised when interpreting this result since another interpretation is possible: farmers did not have time to adopt before the survey and

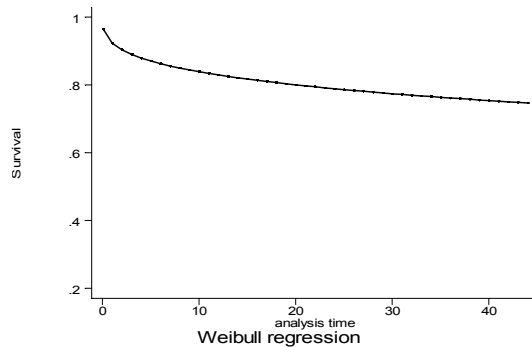
are censored in the analysis. However, an argument in favour of the first interpretation is that the mean (and median) of the adoption duration is less than the duration between 1990 and the year of the survey: the "mean adopter" or the "median farmer" would have had time to adopt.

**Graph 3a: Estimated survival as "non- adopter" before and after 1990**

"Non adoption" distribution before 1990



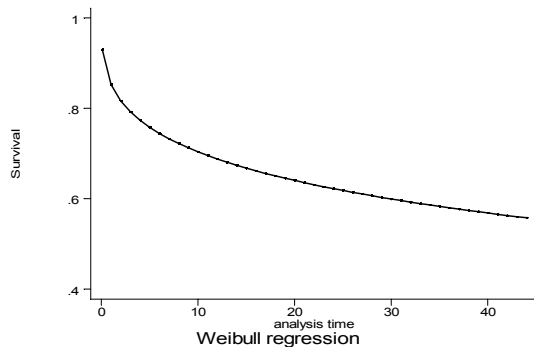
"Non adoption" distribution after 1990



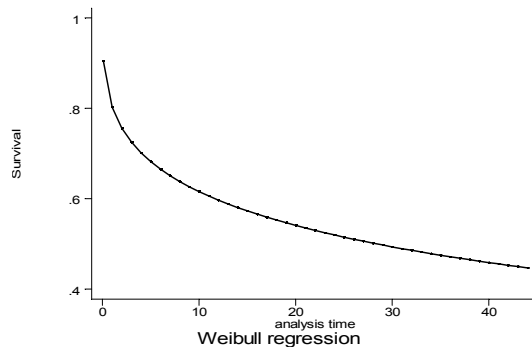
The following graph presents the survival functions when distinguishing the farmers between those with a local cow before the adoption and those without: the result is consistent with the hypothesis that farmers who had some experience in livestock production and/or who had another way of financing the entry cost into dairying adopt faster:

**Graph 3b: Estimated survival as "non- adopter" with and without a local cow**

"Non adoption" distribution for farmers without a local cow before adoption



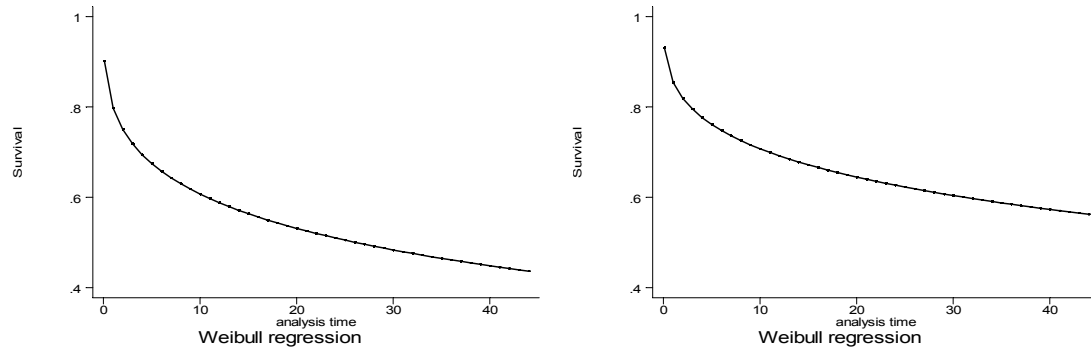
"Non adoption" distribution for farmers with a local cow before adoption



Finally, farmers in tick- infected areas take more time to adopt the grade cattle technology, as hypothesised in the model and as shown on the following graph: the survival of farmers in tick-free area as non- adopters is lower compared to farmers facing harder environmental conditions.

**Graph 3c: Estimated survival as "non- adopter" in tick- free and tick- infected areas**

"Non adoption" distribution for farmers living in tick- free area      "Non adoption" distribution for farmers living in tick- infected area



## **5. Conclusions of chapter 4**

The analysis conducted in this chapter aims at identifying the factors explaining the adoption of grade cattle technology in a dynamic framework that takes into account the accumulation of savings and the "learning- from- others". Moreover, the use of duration models enables to introduce time- series data for key factors: land size, family labour availability and access to credit.

Different estimation methods are conducted: semi- parametric and parametric specifications are estimated and different parametric specifications compared. The results are relatively consistent across the different methods, thus giving confidence in the conclusions that are drawn from the analyses. In fact, the results are consistent between the semi- parametric results (Cox) and the preferred parametric specification (weibull): while access to credit and higher revenues from past agricultural activities accelerate the adoption, results suggest that the speed of adoption has slowed down after the liberalisation of the dairy industry. This result is also consistent with the conclusions derived from the static analysis presented in chapter 3. Another result worth emphasising concerns the "learning- from- others": farmers whose neighbours rear grade cattle adopt faster, reflecting positive externalities in the adoption. The result holds when controlling for the agro- climatic factors that could have explained the observed relationship, had they not been included. Finally, transaction costs are introduced using distances to urban centres

distinguishing between tarmac roads and dry- weather roads. Results show that farmers located 10 km further on a dry-weather roads needs on average almost 11 years more to adopt. The effect is thus substantial and comparison of the estimated marginal effects of the different variables (table 7) show that bad road infrastructures appear to be a major constraint to the adoption of grade cattle technology in the dynamic framework.

When comparing the results of the different parametric specifications, results are consistent except for the access to credit variable that is not significant in the two more flexible specifications (log- normal and log- logistic). Moreover, family labour availability appears to have a significant and positive role in the adoption process in these two specifications: a farmer with more adults on farm is induced to adopt dairying since the activity is labour intensive. However, a likelihood test has shown that the weibull specification is preferred and these two last results have thus to be interpreted with caution.

The three last chapters have dealt with the incentives and constraints to the adoption of the grade cattle technology by Kenyan smallholders. However, the acquisition of a grade animal is only a first step to reach a higher productivity in the dairy enterprise. Among the adopters, levels of milk production vary widely: the analysis of the intensification level, defined as the level of milk produced on farm per unit of land, constitutes the topic of the next chapter.

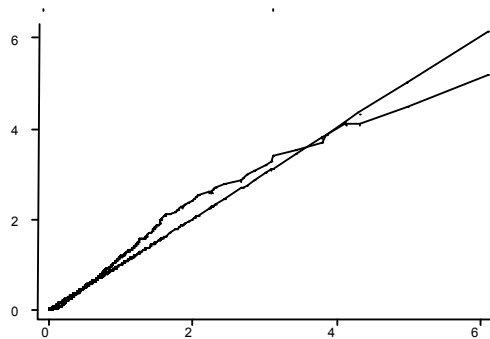
## **Annexes**

### Annex 1: Specification checking of the Cox proportional hazard model using the Cox-Snell residuals

According to Kiefer (1988), residual analysis is the key to assess a specification. If the specification is correct, the "generalised residuals" or Cox Snell residuals should have, in the absence of censoring, a standard exponential distribution with hazard ratio one. The steps are the following:

- compute the Cox Snell residuals defined as the estimated values of the integrated (cumulative) hazard
- consider the residuals as a time variable, censored the same way as the adoption spells, and calculate an empirical estimate of the cumulative hazard function based on the Kaplan- Meier survival estimates
- plot the residuals cumulative hazard function against the residuals: if the specification is correct, the function should agree with the 45° line (which is the actual integrated hazard for the standard (unit) exponential distribution).

The following graph shows the Cox- Snell residuals cumulative hazard function (using the Kaplan- Meier survival estimator): the curve is relatively close to the 45° line at the beginning, then deviates. The results have thus to be interpreted with caution.



Annex 2: Maximum likelihood duration models using four different parametric specifications

Variables	Exponential	Exponential	Weibull	Weibull
<i>Household human capital assets</i>				
age of the household head at establishment	0,001	0,001	-0,007	-0,005
sex of the household head	0,054	0,050	0,053	0,050
years of formal education of the household head	-0,034**	-0,034**	-0,088**	-0,088**
number of adults in the household ♦	0,006	0,007	-0,024	-0,016
dependency ratio ♦	2,054**	1,032***	6,424***	1,663**
<i>Household physical assets</i>				
1 if there was a local cow before adoption	-0,475***	-0,480***	-0,939***	-0,925***
total acreage owned ♦	-0,032***	-0,031***	-0,062***	-0,065***
predicted probability for access to credit ♦	-0,569	-0,837**	-0,375	-1,563*
predicted probability for credit * dependency ratio ♦	-1,939*		-8,321***	
total rainfall since farm establishment ♦	-0,024***	-0,024***	-0,126***	-0,127***
rainfall* dependency ratio ♦	0,000		-0,010	
rainfall* land size ♦	0,000		0,000	
<i>Role of time in adoption</i>				
idiosyncratic human time: year of farm establishment	-0,112***	-0,112***	-0,221***	-0,218***
idiosyncratic human time: dummy variable for “settlers”	-1,086***	-1,077***	-2,486***	-2,421***
pure historical time: dummy variable for liberalisation ♦	1,317***	1,323***	3,273***	3,230***
<i>External factors</i>				
adjusted milk price ♦	-0,548	-0,548	0,184	0,294
adjusted tea price ♦	0,000*	0,000*	0,000	0,000
distance to Nairobi on all weather roads, bound surface	0,017	0,016	0,006	0,002
distance to Nairobi on all weather roads, loose surface	-0,069	-0,067	-0,201	-0,211
distance to Nairobi on dry weather roads	0,387***	0,396***	1,040***	1,078***
distance to two other urban centres on all weather roads	0,095	0,094	0,268*	0,259*
% of "adopters" in the division ♦	-0,013***	-0,013***	-0,037***	-0,036***
availability dairy co-operative in the neighbourhood ♦	0,058	0,061	-0,127	-0,089
presence of the NDDP in the district ♦	-0,231*	-0,245**	-0,521	-0,543*
annual precipitation / potential evapotranspiration ratio	-0,022***	-0,022***	-0,044***	-0,044***
presence of ticks in the area	0,423**	0,419**	1,067***	1,023***
<i>Constant</i>	14,347***	14,504***	27,452***	27,716***
Ancillary parameters			0,34	0,35
Log-likelihood	-1378.313	-1379.737	-1127.941	1132.001
Log-likelihood test	Do not reject the constraints		Reject the constraints at 4%	

Variables	Log- normal	Log- normal	Log- logistic	Log- logistic
<i>Household human capital assets</i>				
age of the household head at establishment	-0,026	-0,025	-0,015	-0,014
sex of the household head	0,070	0,088	0,049	0,057
years of formal education of the household head	-0,151***	-0,152***	-0,120***	-0,123***
number of adults in the household ♦	-0,216**	-0,217**	-0,190**	-0,190**
dependency ratio ♦	4,739**	2,704***	2,682	2,064***
<i>Household physical assets</i>				
1 if there was a local cow before adoption	0,369	0,363	0,234	0,209
total acreage owned ♦	-0,048**	-0,053***	-0,045**	-0,049***
predicted probability for access to credit ♦	-0,099	-0,590	-0,659	-0,826
predicted probability for credit * dependency ratio ♦	-4,285		-1,139	
total rainfall since farm establishment ♦	-0,145***	-0,154***	-0,068***	-0,073***
rainfall* dependency ratio ♦	0,004		0,000	
rainfall* land size ♦	-0,002		-0,001	
<i>Role of time in adoption</i>				
idiosyncratic human time: year of farm establishment	-0,156***	-0,155***	-0,129***	-0,125***
idiosyncratic human time: dummy variable for "settlers"	-2,422***	-2,302**	-2,225**	-2,095**
pure historical time: dummy variable for liberalisation ♦	2,143***	2,104***	1,467***	1,406***
<i>External factors</i>				
adjusted milk price ♦	-2,194	-2,121	-1,342	-1,294
adjusted tea price ♦	0,000	0,000	0,000	0,000
distance to Nairobi on all weather roads, bound surface	0,029	0,030	0,017	0,015
distance to Nairobi on all weather roads, loose surface	0,058	0,073	0,048	0,041
distance to Nairobi on dry weather roads	1,160***	1,129***	0,902***	0,857***
distance to two other urban centres on all weather roads	0,145	0,141	0,069	0,069
% of "adopters" in the division ♦	-0,042***	-0,041***	-0,039***	-0,039***
availability dairy co-operative in the neighbourhood ♦	-0,215	-0,220	-0,185	-0,203
presence of the NDDP in the district ♦	-0,288	-0,247	0,105	0,149
annual precipitation / potential evapotranspiration ratio	-0,086***	-0,086***	-0,067***	-0,066***
presence of ticks in the area	0,387	0,395	0,334	0,346
<i>Constant</i>	27,518***	27,597***	22,391***	22,126***
Ancillary parameters	3,34	3,34	1,58	1,59
Log-likelihood	-1191.359	-1193.074	-1239.822	-1240.702
Log-likelihood test	Do not reject the constraints		Do not reject the constraints	

\*\*\* indicates that the coefficient is statistically significant at 1%, \*\* at 5% and \* at 10%. A positive coefficient indicates that the covariate has a positive effect on the duration before adoption, i.e. "delays" adoption; a negative coefficient indicates that the covariate "accelerates" the adoption.

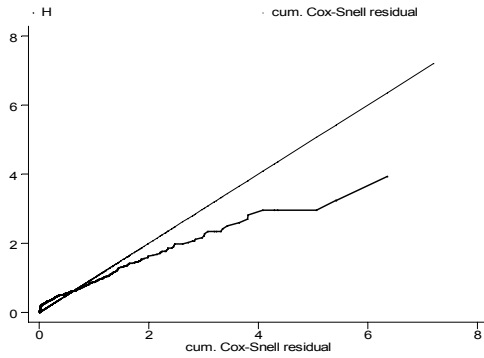
♦ indicates that the variable is a time-variant covariate.

Annex 3: Specification checking of the parametric estimations using the graphical method

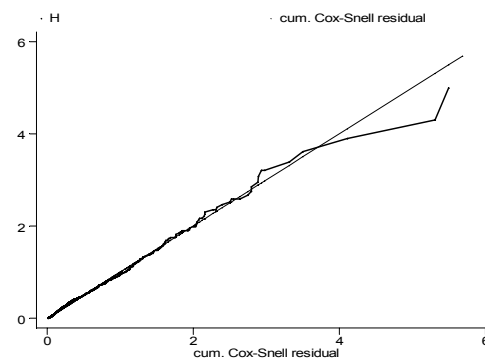
The following graphs present the Cox- Snell residuals cumulative hazard function. The details of the computation are exposed in annex 1 in the case of the Cox semi- parametric estimation. The graphs are computed for the preferred specification (i.e. constraint models for the exponential, log- normal and log- logistic and complete model for the weibull specification).

The curves deviate for the exponential and log- logistic specifications; it can then be concluded that these two specifications dot not fit the data properly. On the contrary, the curves for the weibull and the log- normal specifications are relatively close to the 45° line: results can thus be considered reliable.

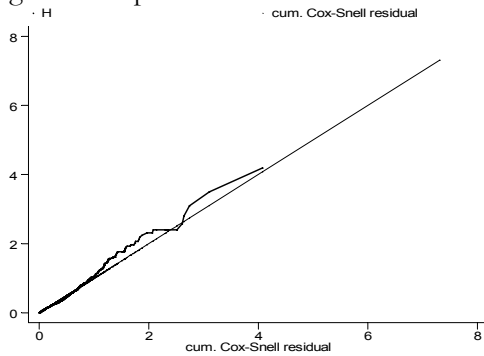
Exponential specification



Weibull specification



Log-normal specification



Log-logistic specification



Annex 4: Discriminating between the four specifications: the Akaike information criterion

When parametric models are nested (e.g. weibull and exponential, for which the weibull parameter is equal to 1), a likelihood ratio test is appropriate. However, it is not when the models are not nested. Akaike proposes comparing the log likelihood obtained for each specification and penalising them by the number of parameters being estimated in a particular model (Statacorp). More specifically, the criterion (AIC) is defined as:

$$AIC = -2 \cdot (\log \text{likelihood}) + 2 \cdot (c+p+1)$$

where  $c$  is the number of model covariates and  $p$  the number of model-specific ancillary parameters.

The preferred model is then the one with the smallest AIC value. The following table presents the results of the test:

	Exponential		Weibull		Log normal		Log logistic	
Log likelihood	-1378,31	-1379,74	-1127,94	-1132,00	-1191,36	-1193,07	-1239,82	-1240,70
AIC	2812,63	2809,47	<b>2313,88</b>	<b>2316,00</b>	2440,72	2438,15	2537,64	2533,40

The weibull specification is thus preferred, since it obtains the smallest AIC value (when comparing the pairs of AIC values – constraint and complete specifications). This result is consistent with the tests based on the Cox-Snell residuals.

## Chapter 5

### Intensifying the dairy activities: who and how?

---

#### **1. Introduction**

The focus of the previous two chapters has been the analysis of the determinants of the adoption of grade cattle technology by smallholders. Grade cows have a higher milk yield potential, but the adoption of grade cattle is only a starting point for a higher productivity on these farms. After getting a grade animal, farmers differ in the level of milk production that is achieved on their farms. Milk production varies widely, even within similar agro- climatic zones.

This chapter aims at getting a better understanding of the observed variability in the intensification levels, where the intensification level is defined as the total milk production given the land availability. Another related issue concerns the strategies adopted by farmers to reach higher levels of intensification. The analysis of the intensification level is crucial as it seeks to identify the possible constraints preventing smallholders from reaching higher productivity in their dairy activities.

Section 2 of this chapter introduces the concept of intensification in dairy farming and explains the data requirements. Because the assessment of the intensification level requires the estimation of the milk production per year, section 3 describes how total milk production is computed using a supply function approach. In section 4, the determinants of the intensification level are analysed by looking at the factors discriminating farmers with different levels of intensification; in other words, the analysis seeks to answer the question: who intensifies? Section 5 deepens the analysis and identifies the different ways to intensify, answering the question: how to intensify? Two routes towards intensification are identified: a farmer intensifies through higher milk yields (high milk production per cow) or through more cows per unit of land. The last section concludes on the constraints faced by smallholders when intensifying their dairy activities.

## **2. Intensification level in dairy farming: definitions and issues**

### 2.1. Definitions of intensification level in dairy farming

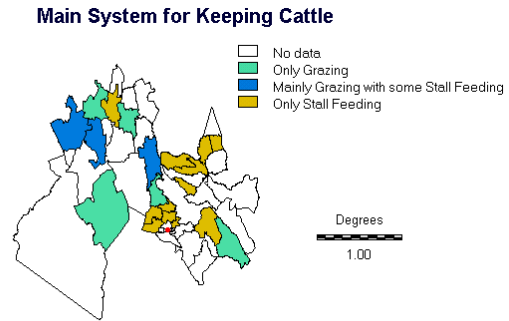
Due to the land constraint in the East Africa Highlands, smallholders' increasing objective is to intensify the agricultural activities (Carr 1989). In the particular case of the Kenyan Highlands, smallholders dairy systems are currently intensifying, i.e. farmers reach higher level of milk production per unit of land through an increase in the use of inputs. Inputs are the type of animals (from indigenous cows to crossbred or grade cows as described in chapters 3 and 4); capital (from open grazing to a "zero-grazing" unit); labour (from herding to labour-intensive activities, e.g. collecting and preparing feed, feeding the animals); and other external inputs, especially feeds (concentrates and other high-quality feeds) and veterinary services.

Agricultural intensification is generally defined as productivity per unit of land (Boserup 1990). In the case of dairy farming, the same indicator is used: intensification is defined as the quantity of milk produced per land acreage. Of note is the fact that the major constraint in livestock production identified in the literature is the feed availability; it would have then been relevant to define intensification in dairy farming with respect to feed resources. However, there is no uniform and practical unit for feed: intensification in dairy farming is thus defined in the subsequent analyses with respect to land resources (Staal 1999).

Besides this indicator, intensification in dairy farming can be measured by "secondary" indicators that capture the input use on dairy farms (Baltenweck *et al.* 1998). Although such indicators provide interesting insights on the process at play, they do not offer a complete conclusion and farmers' classification into different intensification level classes may not coincide across indicators. By consequence, the analysis conducted in this paper focuses on the main indicator that offers an unique measure of the intensification level. It is however worth presenting succinctly some of these "secondary" indicators. Averages by district are calculated and mapped in order to visualise the patterns of intensification using these different indicators.

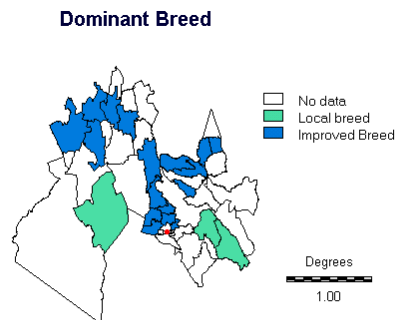
They are the following:

1. The dominant system for keeping cattle: four systems are identified: "only stall feeding", "mainly stall feeding with some grazing", "mainly grazing with some stall feeding" and "grazing". Using this indicator, most intensified farms are defined as those using the "only stall feeding" system. The map shows that this system is the most observed one in the surveyed area (12 over the 20 divisions).



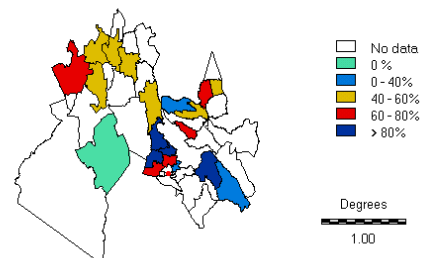
Of note is the fact that divisions where stall feeding is dominant are the areas close to Nairobi whose position is described by the red circle.

2. The dominant breed: the distinction is between local breed cattle and grade cattle. The dominant breed differs in a way similar to the system for keeping cattle. In fact, grade animals are more present where the main system is stall feeding while local animals are found in grazing areas.



3. The percentage of cost of purchased feed in total feeding cost, where the total cost is equal to the cost of purchased feed plus the opportunity cost of growing feed on-farm<sup>15</sup>. The higher the indicator, the higher the intensification level. Using this indicator, farms close to Nairobi seem to be more intensified, although the pattern reveals exceptions, e.g. Molo division at the upper left (in red).

**Percentage of Cost of Purchased Feed in Total Feeding Cost**



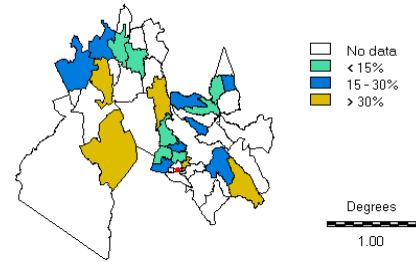
<sup>15</sup> The opportunity cost of growing feed on-farm is the reported rental value of the land planted in *Napier* (fodder) and in pasture. Of note is the fact that this method of calculation makes the opportunity cost of growing feed on-farm very high for farmers in Nairobi district since rental value of land is boosted by population pressure.

4. The percentage of farmers hiring labour for dairy activities<sup>16</sup>:

since intensification needs the use of more inputs including labour, a farmer is classified as more intensified if she hires external labourers to work on dairying. However, this indicator does not provide clear-cut insights since a high level can be explained by two very

different ways: for high-intensified system, labour is necessary to "cut- and- carry" fodder to the animals while some farmers hire labour to graze the animals in the less-intensified systems (e.g. Narok district, the area at the lower left).

Percentage of Households Hiring Labour for Dairy Activities



2.2. Analysis of the intensification level

The literature on intensification originates with Boserup. In her model, intensification (defined as yield increase) is described as an endogenous process in response to increased population pressure. As the ratio land to population decreases, the use of labour increases resulting in diminishing returns. In order to produce enough, farmers are thus induced to adopt a technological innovation. For example, land is cropped more intensively through diminishing fallow periods and introduction of the plough. Yields per unit of land increase but the labour productivity (output per unit of labour) may decline or stagnate. Despite the simplicity of the model, Cuffaro (1997) shows that aggregate data for the period 1962-92 for Developing countries, including Kenya, are consistent with Boserup's model. These countries meet the hypothesis of an increase of the population pressure on land and as Boserup's model predicts, the land productivity increased but the growth of the labour productivity has been much lower or has declined. Using macro- economic data, Islam and Taslim (1996) analyse the adoption of high-yielding varieties seeds in Bangladesh. They show that the adoption of this new technology was driven by demographic pressure, as Boserup had postulated. However, the production increase due to the technological change did not entail increased labour productivity.

However, Boserup's model does not address some critical issues. Concerning the process of adoption, the focus is on technology transmission with no insights on the origin of the

---

<sup>16</sup> or more precisely the percentage of farmers hiring external labour whose work is "mainly" related to dairy activities. "Mainly" means that they spend more than 50% of their time on dairy activities for all the districts, except for Kiambu where it means that it is the labourer who is primarily responsible for carrying out the dairy activities.

technology. Another shortcoming deals with the conditions for the technology adoption: there may be some constraints to adoption, i.e. lack of capital (as showed in chapters 3 and 4) and non-availability of inputs. Another limitations of the model is emphasised by Cuffaro: the model does not take into account the questions of distribution and entitlements since it postulates that increasing food needs induce endogenously the adoption of a new technology to increase food production for the whole community. There is however a certain heterogeneity across farmers, raising the need for a micro- analysis that follows in the subsequent sections.

In response to these limitations, Lele and Stone (1989) distinguish two types of intensification:

- the first type, identified by Boserup, occurs spontaneously as land is cropped more frequently in response to higher population densities: it is the “autonomous intensification”. The crucial factor is the ratio of land to labour. Lele and Stone note that Boserup's view of the intensification process is consistent with Hayami and Ruttan's "induced innovation" theory (Hayami and Ruttan 1984 and 1985): land- saving technologies appear in countries with an inelastic land supply while mechanical technologies are adopted in areas with inelastic labour supply.

- the second type depends on policy and incentives for a shift to crops of higher value or higher yields, or to more productive land: it is the “policy-led intensification”. The factor explaining intensification is the availability of a suitable and affordable technology. The same line of argument is used by Smith *et al.* (1994) in a study concerning the introduction of an improved maize variety in Nigeria. The technology introduced provided farmers with a profitable crop given the area ecological conditions; moreover, preconditions for the adoption, i.e. a good road system and availability of extension services were met.

These considerations will be introduced when determining the explanatory factors of the intensification level: it is the focus of section 4.

As explained above, intensification level is defined in this analysis as the quantity of milk produced relative to the land acreage. This indicator thus requires calculating the total milk production per year and per household. The calculation of milk production is complicated by the fact that the production is daily (while the cow is in milk) and not constant over time. In fact, milk production is shown to decline after calving. Daily data on milk production are not available, thus the need to first estimate econometrically a production function in order to predict the level of milk production. This is the focus of section 3.

### **3. Milk production: a production function approach**

In order to compute the level of milk production per farm per year, two steps are necessary that correspond to the following two sections. The first section deals with the econometric estimation of the milk production at some specific dates; the second section presents the method for predicting the level of milk production per year per household based on the econometric results of the first section.

#### 3.1. Estimating milk production per day

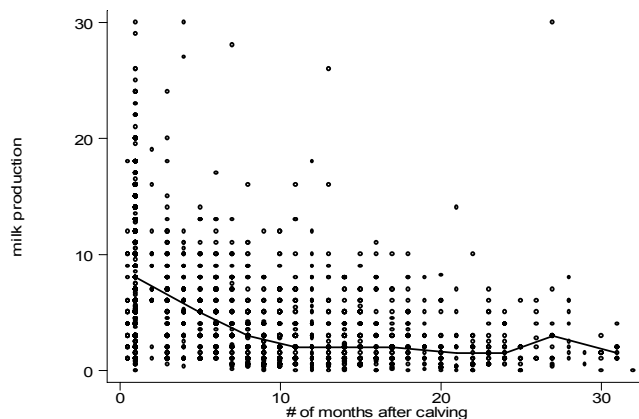
Data on milk production were collected using recall questions at three points of the lactation stage: at calving, at the end of the lactation, and at the time of the survey. Data on the second last and last calving dates were collected, as well as the last date when the farmer stopped milking the cow (end of the lactation). Ideally, there are thus three observations per cow, which are used to estimate the milk production using a production function approach (by opposition to a supply function). A total number of 2629 observations are available to estimate the parameters of the production function. These observations correspond to 1424 cows, present on 812 farms.

The determinants introduced in the analysis are classified in three levels: the month of lactation at time  $t$  for the cow  $i$  on the farm  $j$  (in log form); other characteristics of the cow  $i$ ; some characteristics of the farm  $j$  and the environment. The equation is written as:

$$Y_{i,j,t} = a. \ln(\text{month}_t) + b'. C_{ij} + c'. F_j + d'. E_{ij}$$

where  $i$  represents the cow,  $j$  the farmer and  $t$  the month of lactation.  $Y_{i,j,t}$  is the daily milk production of the cow  $i$  in the farm  $j$  after  $t$  months after calving.  $C_{ij}$  is a vector of cow characteristics,  $F_j$  a vector of farmer's characteristics and  $E_{ij}$  represents the environment characteristics.

The month of lactation is introduced because milk production peaks shortly after calving and then declines (for tropical dairy cattle, milk production peaks at 4 to 8 weeks after calving). Following Omore (1997), a log form of the number of month after calving is preferred. The following graph presents the observed milk production levels at the corresponding months after calving. Despite the presence of outliers, graph 1 shows clearly that milk production declines progressively after calving.

**Graph 1: Observed milk production levels at corresponding months after calving**

Cow characteristics that are introduced are the cow genotype (whether a local cow or a high-grade cow, by comparison with a crossbred cow) and the number of calving. A local cow is supposed to have a lower milk production, *ceteris paribus* compared to high-grade cow. Milk production is increasing in the number of previous calvings, up to a certain point from which the production is declining (Essl 1998): the non-linearity is captured by introducing the variable “number of previous calvings” and the variable squared.

Farm characteristics that are introduced are production characteristics that are the same for each cow. Feed availability is an important determinant of milk production, therefore on-farm planted fodder and purchased feed are introduced. Purchased feeds are introduced through the total feed expenditures per tropical livestock unit (TLU<sup>17</sup>) on farm. Expenditures are preferred to quantities since it is straightforward to sum them (without using conversion factors) and because the price is assumed to reflect the quality of the feed (with the hypothesis that high-quality feeds are more expensive). On-farm planted fodder is included in the estimation by using the acreage planted in Napier grass (main fodder in the studied area), maize acreage and land under pasture. Maize acreage is introduced because feeding the animals with maize leaves (thinning) and maize stovers (green and dry) is a common practice. Two other farm characteristics are introduced: because a farmer practising zero-grazing is expected to have a higher production (because of better management and less risks to contract tick-borne diseases due to the confinement), a dummy variable indicating whether the animals are stall-fed is introduced. Finally, a dummy variable for vaccination is introduced, reflecting the availability of veterinary services in the area and the general health conditions of the animals.

<sup>17</sup> Tropical livestock unit is defined by applying the following weights to the different categories of animals: 1 for bulls and castrated males, 0.7 for cows, 0.5 for immature males and heifers and 0.2 for calves (male and female).

Environment characteristics are introduced through a climatic variable that reflect the fodder availability in the area (annual PPE- annual precipitation over overall potential evapo-transpiration ratio described in chapter 1).

Ideally, there are three observations per cow (at calving, at the time of the survey and at the end of the lactation). Observations from the same cow cannot be considered independent and the econometric methods that are conducted control for it by using a "robust" (or White) estimator of variance. It would have been adequate to control as well for the fact that observations for cows belonging to the same farmer are not independent, but two controls are not possible.

Two main specifications are conducted. The first specification introduces separately the acreages planted with Napier, maize and under pasture. The second specification introduces on-farm feed availability as a weighted average of the acreages under Napier, maize and pasture. The weights correspond to the expected dry-matter yield of the specific type of fodder<sup>18</sup>. The advantage of the formulation is to compute a unique indicator of the on-farm feed availability, but at the cost of hypotheses on dry-matter yields.

Because the presence of outliers could not be ruled out, robust regression methods are conducted in order to reduce the possible influence of these observations on the results. Three different methods are used; details are presented in annex 1. The first method is the classical weighted least squares method that controls also for the non-independence between observations from the same cow. The second method is a robust regression, whereby gross outliers are detected and their influence on the estimated coefficients decreased using weights. The third method is the median regression: the objective is to estimate the median of the dependent variable (and not the mean), conditional on the values of the independent variables. Median regression methods are less sensitive to outliers. However, this method seems to understate the standard errors for heteroscedastic errors, a bootstrapping method is then used in order to estimate the standard errors.

Results from the two specifications and using the three methods are presented in annex 2. Table 1 presents the results of the estimation of the first specification (introducing separately the three

---

<sup>18</sup> Napier is hypothesised to produce 8 tons of dry-matter by acre, maize 3 tons and pasture 2 tons (I.L.R.I. staff estimations).

types of fodder sources) using the weighted least squares estimation method; milk production levels are then computed using those results.

**Table 1: Results of the regression of milk production: WLS**

Milk production per day at specific dates	Coefficient	T-Statistics	Level of significativity
<i>Cow characteristics</i>			
Ln(month of lactation)	-1.962	-33.677	0.000
1 if the cow is a local breed cow	-2.061	-6.159	0.000
1 if the cow is a high-grade cow	0.540	2.066	0.039
Number of previous calvings	0.520	4.225	0.000
Squared value of number of calvings	-0.040	-3.172	0.002
<i>Farm characteristics</i>			
1 if the cow is stall-fed	0.810	3.084	0.002
1 if the cow is vaccinated	0.357	1.515	0.130
Acreage in maize per TLU	0.019	0.137	0.891
Acreage in Napier per TLU	0.048	0.173	0.862
Acreage in pasture per TLU	0.139	1.285	0.199
Feed expenditures per TLU	6.10 <sup>5</sup>	1.138	0.255
<i>Environment characteristics</i>			
Annual PPE (*100)	0.069	7.545	0.000
<i>Constant</i>	1.172	1.608	0.108
<i>Number of observations</i>	2629		
<i>R-squared</i>	0.360		
<i>F(12, 1324)</i>	155.17		

Results show that milk production declines with time after calving, as the coefficient of the month of lactation shows: since  $\ln(1)$  equal 0, milk production is maximum during the first month after calving. After six months, milk production has declined on average by 3.5 litres, *ceteris paribus*. Dummy variables for the genotype have the expected sign: local (high-grade) cows have a lower (higher) milk production. The result on the number of calvings shows that milk production is high in the early stages of the cow production life, then declines after 6.4 calvings on average. This result is consistent with the literature (Essl 1998).

Farm characteristics proved to be important factors. Stall-fed cows have higher production levels. Results using robust regression methods show that feed expenditures per TLU have a positive effect on milk production: this result is consistent with the view that concentrates (principal type of purchased feed) have a positive effect on milk production. A similar result is obtained by Schaik *et al.* (1996) in a study of eighteen farm in Murang'a district, Kenya (one of the surveyed area), using a Monte- Carlo simulation to evaluate the financial and economic impacts of technology changes on a dairy farm. The authors show that the amount of concentrates fed influences positively the level of milk production (as well as calving intervals).

However, quantity of fodder planted on-farm has a positive and expected effect only in some specifications. This can be related to the results of other studies concluding that feed offered to cattle is too low on smallholder farms in Kenya to reach high levels of milk production. In fact, the quantity of feed offered to the animals is only enough to maintain the body weight but not to achieve high levels of milk production (Omoro *et al.* 1999). In some cases, feed offered is not enough to reach maintenance levels and the animals lose weight, especially during dry seasons (Utiger, personal communication).

### 3.2. Predicting milk production per year per household

In the previous section, milk production is estimated at some specific dates, i.e. at calving, at the end of the lactation and at an intermediate date (when the survey was conducted). The next step is to calculate the milk production per lactation per cow: it is equal to the sum of the monthly levels of milk production during the lactation period.

Mathematically:

$$\begin{aligned}
 Y_{i,j, \text{lactation}} &= 30. (Y_{i,j,1} + Y_{i,j,2} + \dots + Y_{i,j,n}) \text{ where } n \text{ is the lactation length (in month)} \\
 &= 30. (a. \ln(1) + cste_{i,j}) + 30. (a. \ln(2) + cste_{i,j}) + \dots + 30. (a. \ln(n) + cste_{i,j}) \\
 &\text{where } cste_{i,j} = b'. C_{i,j} + c'. F_j + d'. E_{i,j} \\
 &= 30. (a. \text{sum}(\ln(t)) + n. cste_{i,j}) \text{ where the sum runs from 1 to } n
 \end{aligned}$$

The coefficients  $a$ ,  $b'$ ,  $c'$  and  $d'$  were estimated in the previous step. Milk production per year is then computed by dividing the milk production per lactation by the calving interval (to get an average per month) and multiply by 12.

The last step is to calculate the milk production per year per household, by summing the milk production per cow for all the cows on farm. Because of missing data (especially on dates), milk production could only be predicted for 198 households (corresponding to 477 cows). Moreover intensification level is only assessed for farmers with crossbred and/or high-grade cows; in addition, the lower 5% and higher 5% values are excluded in order to exclude outliers from the analysis. A total of 165 households are finally kept for the intensification level analysis.

Based on these estimates, indicators for the dairy activities are calculated and are presented in table 2 below.

**Table 2: Using the results from the milk production function to estimate some production parameters**

Variable	Number of obs.	Mean	Standard deviation	Minimum value	Maximum value
Milk production per cow per year (litres)	402	1351	658	81	2467
Calving interval (months)	649	18	6	11	38
Lactation length (months)	376	14	5.3	7	31
Fodder acreage per TLU	787	0.47	0.62	0.00	5.08

The average milk production per cow per year is 1351 litres, which is consistent with observed milk yields in Kenya (Ngugi *et al.* 1990). Other interesting production indicators are the calving interval (length between two calvings), the lactation length and fodder availability per animal. The average calving interval approximates 18 months: this figure is above the recommendations (i.e. 12 months) as widely noted in the literature.

Fodder availability is calculated by TLU: the recommendations are 0.51 acre of Napier per TLU unit (KARI Kiambu recommendations) and the calculations show that the average on the surveyed farms is 0.47 acre, thus slightly lower than the recommendation level. However, the large standard deviation shows that there is a large variability across farmers.

#### **4. Determinants of the intensification level**

The last section established the “technical” relationship between the production factors (cow characteristics and inputs) and the level of milk produced by a cow. In this section, the analysis follows a supply approach, whereby the objective of the analysis is to understand the reasons driving intensification. The first sub-section presents the results of the theoretical model of chapter 3 applied to the level of intensification. The econometric method and the empirical results are presented in the second sub-section. The third sub-section analyses the results and concludes.

##### 4.1. The theoretical modelling

Determinants of the intensification level are derived from the model developed in chapter 3. In chapter 3, the reduced-form of the household demand for input cattle was derived; for this analysis the model is used to identify the reduced form for the intensification level, i.e. milk production per unit of land. It can be shown that the reduced form is the same for the output (milk). The only difference deals with access to credit. Since it was hypothesised that access to credit enables relaxing a possible cash constraint when purchasing a grade cow, it only influences the entry cost into dairying; therefore, this variable does not appear among the determinants of

the intensification level. In fact, the decision on the intensification level is taken after the adoption decision. The determinants of intensification are summarised in table 3.

The literature on intensification in section 2 identified two types of intensification. The first type is an autonomous process due to pressure on land: the empirical analysis takes it into account through the variables land size and number of adults on farm. The other type of intensification is policy-led: dairy activities in Kenya have been encouraged by the government through the provision of services, e.g. artificial insemination, veterinary and extension services. The analysis introduces the availability of artificial insemination services to control for this policy. It is not possible to introduce separately the different services since they are often offered as a "package".

**Table 3: Categories of variables**

Theoretical model	Variable in the equation
Household characteristics	
- Household head characteristics	$Z_t^h$
- Other household characteristics	$\bar{H}_t, L_t^m, L_t^w, Z_t^b, \sigma^2$
Prices and transaction costs	
- Milk price and transaction costs	$p_{m,t}, TC_m$
- Wage rate & supervision costs	$w_t^m, w_t^w, s$
- Other prices	$p_{c,t}, p_{j,t}, p_{v,t}$
Area characteristics	
- Specific to dairying	$Z_{c,t}^r$
- Both for dairying and crop activities	$Z_{c,t}^r, Z_{j,t}^r$

When comparing with the adoption analysis, it can be noticed that the computation of some variables differs between the two estimations. While the analysis of the adoption decision required evaluating some variables at the time of the farm establishment, variables introduced for the analysis of the intensification level should all be evaluated at the time of the survey. In fact, the observed dependent variable is the milk produced the year of the survey. Variables that are not computed the same way for the two analyses are the land acreage and milk price.

## 4.2. The empirical analysis

### 4.2.1. The econometric issue

The first decision taken by a farmer is whether to adopt a grade cow. She then takes a decision on the level of intensification. Because there may be unobservable factors at work, the two decisions (adoption and level of intensification) cannot be analysed separately. More specifically,

the hypothesis that high skilled farmers both adopt and reach high levels of intensification cannot be ruled out. However, skills are unobservable characteristics that can only be approximated by variables like the education level. The issue is then to deal with a selectivity bias whereby the level of intensification is only observed for some categories of farmers.

The classical way to deal with selectivity bias is to compute the Inverse Mills Ratio (IMR) using the probit of the adoption analysis. They are then introduced among the explanatory variables in the equation of intensification level. It is the Heckman method. The details of the method and the computation of the IMR are explained in annex 3.

Nevertheless, this method relies on the strong assumption that the residuals follow a normal distribution. Moreover, the Heckman procedure gives rise to heteroscedasticity (Greene 1997, p. 977). According to Deaton (1997, p.104), the other issue when using this method is the problem of identification: there should be at least one variable in the switching regression (i.e. adoption decision) that does not appear in the substantive equation (i.e. intensification level); and this variable should be continuous. In this analysis, the variable access to credit, which is a continuous variable (predicted probability of access to credit), is a determinant of adoption but not of the intensification level. Moreover, some variables are not computed the same way for the two equations: the variables “land acreage” and “milk price” are evaluated at the time of establishment in the adoption equation and at the time of the survey in the intensification level equation. The requirements to use the Heckman method are thus met.

However, because the Heckman method relies on the normality of the residuals, other methods to deal with the selectivity bias are used. Deaton (p. 105) proposes a “mixture of parametric and nonparametric techniques” to control for the non-observable factors that may be at play in both equations. Instead of introducing the IMR, a polynomial form of the estimated index of adoption or of the estimated probabilities of adoption is introduced. Annex 3 presents the details of the method.

The three methods are used. In fact, comparing the results obtained by the different methods gives more reliance in the obtained results. The Heckman method is theoretically very strong but requires accepting the normality assumption. The two other methods do not postulate any assumptions on the distribution of the residuals but lack a theoretical background; they are more intuitive.

4.2.2. The empirical estimation

As in the case of the estimation of the milk production function, different regression methods are used. The first method is a weighted least squares regression; the second method is a robust regression while the third is a median regression. Details on the regression methods are presented in annex 1.

The results are presented in annex 4.1. There are nine regressions: three regression methods and three methods to control for the possible selectivity bias. In table 4 below are presented the results of the weighted least squares regression using the IMR.

On overall, goodness- of- fit indicators are relatively low: the R-squared indicators range from 0.23 to 0.38. However, the following comments can be derived from the analysis of the table in annex 4.1. and the table 4 below.

**Table 4: Results of the regression of intensification level: WLS**

The dependent variable is milk production per land size			
Variables	Model variables	Coefficients	Significativity
<i>Household &amp; household head characteristics</i>			
- age of the household head	$Z_t^b$	-14,449	0,009
- sex of the household head	$Z_t^b$	-102,999	0,499
- education level (years)	$Z_t^b$	24,911	0,084
- 1 if kikuyu	$Z_t^b$	169,545	0,604
- land acreage	$\overline{H}_t$	-16,376	0,002
- number of adults	$L_t^m, L_t^w$	-11,965	0,670
- proportion of female adults	$L_t^m, L_t^w$	570,106	0,157
<i>Prices and transaction costs</i>			
- milk price in the sublocation	$p_{m,t}$	23,293	0,564
- distance on all-weather roads	$TC_m$	-13,783	0,054
- distance on dry-weather roads	$TC_m$	2,141	0,904
- wage rate	$w_t^m, w_t^w$	-2,809	0,210
<i>Area characteristics</i>			
- dairy cooperative availability	$TC_m, Z_c^r$	327,595	0,057
- extent of A.I. availability	$Z_{c,t}^r$	151,417	0,465
- 1 if it is a tick-infected area	$Z_{c,t}^r$	162,140	0,342
- annual PPE (*100)	$Z_{c,t}^r, Z_{f,t}^r$	28,577	0,001
- minimum temperature	$Z_{c,t}^r, Z_{f,t}^r$	51,281	0,176
<i>Constant</i>		-1558,260	0,215
<i>Mills's ratio</i>		695,342	0,039
<i>Number of observations</i>	159		
<i>R-squared</i>	0.35		

The agro- climatic factors are the only determinant of the intensification level that is significant in all the equations: areas benefiting from better climatic conditions record higher intensification

levels. This result suggests that agro-climatic conditions influence the dairy activities after controlling for the production factors, as fodder availability is higher in areas with better agro-climatic conditions.

Total land size is the other variable with a significant effect on the intensification level: the coefficient is significant in five over the nine regressions. The larger the farm size, the lower the intensification level. This is consistent with the hypothesis that large farms have less incentive to intensify since the land availability is higher. A rather surprising result concerns the family labour availability. The variable has a significant effect in any of the regressions, yet it was expected that it would encourage the intensification level in dairy farming.

Other household variables have an influence on the intensification level: more educated farmers reach higher intensification level on their farms. This result is consistent with the results obtained in the previous chapters: education enables farmers to reach higher productivity levels. On the other hand, farms headed by an older head record lower intensification levels.

Two other variables have a significant effect, but only in few specifications: farms located further away from urban centres have lower intensification levels (in four regressions over nine). This result is expected since one driving force behind intensification is the higher milk price paid in consumer centres; moreover, the proximity with urban centres eases milk marketing since farmers are less dependent on unreliable "hawkers". And availability of dairy co-operatives fosters intensification also (in three regressions over nine). This result is consistent with the hypothesis that dairy co-operatives provide a reliable outlet for milk production, as well as other services (feeds on credit for some of them): farmers are thus induced to increase their dairy activities.

Finally, the control for the selectivity bias proved to be relevant: in fact, the IMR are significant in all the cases. Farmers who adopted grade cattle technology are those who reach higher intensification in dairy farming. However, the two other methods do not present evidence of a selectivity bias: results using the IMR have thus to be interpreted with caution.

#### 4.3. Analysis of the results and conclusions

The analysis conducted in this section aims at identifying who reaches high levels of intensification in dairy farming. Farmers living in areas suitable for dairy farming and those with limited land size reach higher levels of milk production given their land constraint, *ceteris paribus*. However, caution must be exercised when interpreting the results because of the relatively low levels of goodness of fit obtained in the regressions. Because the level of milk production had to be estimated in a first step, the dependent variable is measured with error; it is thus not surprising

to get relatively poor results. Moreover, the intensification level is a “combined” indicator that does not distinguish between the different strategies pursued by farmers. In fact, a same level of the indicator can be achieved through different strategies. The next section aims at solving this shortcoming by using secondary indicators describing the strategies that farmers use in order to increase the productivity of dairy farming.

## **5. The routes toward intensification**

The analysis of the determinants of the intensification level has shown that the major factors driving intensification at the household level are the agro- climatic conditions and the quantity of land owned by the farmer. After answering "who intensifies?", the next step is to analyse the different routes toward intensification, i.e. how do farmers reach a higher milk production, given their land constraint? The objective of this section is to answer this question.

### **5.1. Identifying the routes toward intensification**

Two routes towards dairying are identified: a farmer intensifies through higher milk yields (milk production per cow) or through more cows per unit of land. In fact, the intensification level indicator defined previously can be broken down into those two secondary indicators since

$$\frac{\text{milk production}}{\text{land acreage}} = \frac{\text{milk production}}{\text{number of cows}} \cdot \frac{\text{number of cows}}{\text{land acreage}}$$

The two ways then defined four possibilities:

1. The most intensified systems: high milk yield and high number of cows per unit of land. Farmers in this category have high milk production per cow while at the same time keep many cows, supplementing them with high-quality feeds in order to get above average milk yields.
2. Intermediate level of intensification through on-farm feeding: high milk yield and low number of cows per unit of land. Because the farmer keeps (relatively) few cows, the quantity of on-farm feed available per cow is enough to get high yields.
3. Intermediate level of intensification through “overgrazing”: low milk yield and high number of cows per unit of land. High levels of milk production at the farm level are reached by keeping many cows, each producing little milk.
4. The least intensified systems: low milk yield and low number of cows per unit of land.

It should be noted that the indicator of pressure of land only takes into account the number of cows on farm, as opposed to the total number of cattle. However, when classifying the households according to the total number of cattle on farm (in TLU), 149 households over 164 (90.9%) are classified the same way as when using the number of cows. It is thus assumed that the indicator number of cows per acreage is a relevant indicator.

## 5.2. The empirical analysis

### 5.2.1. Looking at the routes toward intensification

Table 5 presents the means of the determinants of the intensification level when the farmers are classified into four categories, based on the observed median values of the two indicators. A farmer is classified as “most intensified” (“least intensified”) if the ratio of milk production per cow is higher (lower) than the median and the ratio of number of cows per unit of land is higher (lower) than the median. The two other categories are computed in a similar way.

**Table 5: Means of variables, by intensification categories**

Variables	Least intensified	Intermediate through overgrazing	Intermediate through on-farm feeding	Most intensified
<i>Codes for categories</i>	(a)	(b)	(c)	(d)
<i>Number of observations</i>	45	37	33	49
<i>Household &amp; household head characteristics</i>				
- age of the head	57.16 (b,d)	47.68 (a,c)	56.03 (b,d)	48.08 (a,c)
- sex of the head	0.76	0.70	0.82	0.78
- education level (years)	6.80 (b,d)	9.14 (a,c)	7.36 (b)	8.49 (a)
- 1 if kikuyu	0.71 (b,c,d)	0.86 (a,d)	0.94 (a)	1.00 (a,b)
- land acreage	16.06 (b,d)	2.89 (a,c)	14.06 (b,d)	2.44 (a,c)
- number of adults	4.2 (d)	3.78	4.48 (d)	3.59 (a,c)
-proportion of female adults	0.53	0.55 (c)	0.48 (b)	0.52
<i>Prices and transaction costs</i>				
- milk price	16.42 (c,d)	15.79 (c,d)	14.33 (a,b,d)	14.68 (a,b,c)
- distance on all-weather roads	29.02 (c,d)	26.83 (c,d)	33.53 (a,b)	34.87 (a,b)
- distance on dry-weather roads	2.83	2.05	2.54	2.28
- Wage rate	86.38 (b,c)	93.17 (a)	98.06 (a,d)	88.07 (c)
<i>Area characteristics</i>				
- dairy cooperative availability	0.42 (d)	0.46 (d)	0.54	0.63 (a,b)
- extent of A.I. availability	0.23 (d)	0.30	0.28	0.38 (a)
- 1 if tick-infected area	0.78 (b)	0.89 (a,c)	0.70 (b,d)	0.88 (c)
- annual PPE	68.41 (c,d)	69.23 (c,d)	84.94 (a,b,d)	81.30 (a,b,c)
- minimum temperature	9.74 (c)	9.63	8.91 (a,d)	9.70 (c)

Note: The letters into brackets indicate that the mean for this category differs significantly from the mean of the category into bracket (at least at 10%).

The analysis of table 5 identifies three sets of factors explaining the choice of routes: the farm characteristics, the availability of services in the neighbourhood and the agro-climatic factors.

Farm characteristics that differ significantly between categories are the land size, the number of adults and the education level of the household head. As observed in section 4., land size is a major factor driven intensification. Households with limited land sizes are the most intensified and these with high ratios of number of cows per acreage. Farmers with large land size choose the other route toward intensification, i.e. intensification through on-farm feeding. Note that the results are consistent: farmers with the smallest land size are the most intensified while those with the largest land size are the least intensified. The other result deals with the number of adults on farm: strangely, the least intensified farms are characterised by a higher number of adults compared to the most intensified farms. However, the farms with the highest availability of family labour reach intermediate levels of intensification through on- farm feed, a result which is consistent with the observation that labour is needed to prepare the feeds for the animals. The last household characteristic is the head's education level: the least intensified farms are those headed by the less educated farmers: this result is comparable to the result obtained in the intensification level analysis whereby education seems to facilitate reaching higher levels of intensification.

Availability of services seems as well to play a role in the choice of the routes toward intensification: services (existence of a dairy cooperative and availability of artificial insemination services) are more available to the most intensified and those who intensify through “on farm-feeding”. Availability of services could thus be a way to promote intensification and higher milk yields.

Finally, the annual PPE seems to play an important role: the most intensified households and those who rely on on-farm feeds for intensification benefit from better climatic conditions.

### 5.2.2. The econometric analysis

Because a statistical analysis cannot control for the interactions between variables, an econometric analysis is conducted for the two routes identified above: number of cows per unit of land and milk yield per cow. Since it cannot be ruled out that these two decisions are taken simultaneously, a system of equations is estimated and presented in annex 4.2. However, results show that the coefficient of correlation between the residues of the equations is not statistically significant. Estimating independently the two routes will thus not bias significantly the results. Estimating separately the two routes allows using robust methods to gross outliers in the same

way as for the estimation of the intensification level. Results using the different regression methods are presented in table 6 below and in annexes 4.3 and 4.4.

**Table 6: results of the econometric analysis for the two routes for intensification (weighted least squares regressions using the IMR to control for the selectivity bias)**

Variables	Number of cows per land size	Milk yield per cow
	Coefficients & level of significance	Coefficients & level of significance
<i>Household &amp; household head characteristics</i>		
- age of the household head	-0,008***	-0,110
- sex of the household head	0,043	-18,986
- education level (years)	0,011*	6,316
- 1 if kikuyu	0,355***	105,235
- land acreage	-0,007***	-3,593
- number of adults	-0,017	0,301
- proportion of female adults	0,169	-198,881
<i>Prices and transaction costs</i>		
- milk price in the sublocation	-0,003	-10,131
- distance on all-weather roads	-0,012***	-3,506
- distance on dry-weather roads	0,001	4,596
- wage rate	-0,001	0,568
<i>Area characteristics</i>		
- dairy cooperative availability	0,145***	106,464
- extent of A.I. availability	0,213***	114,625
- 1 if it is a tick-infected area	-0,068	153,779*
- annual PPE	0,010**	26,725***
- minimum temperature	0,027	-3,555
<i>IMR for adoption of grade cows</i>	0,637***	138,407
<i>Constant</i>	-0,235	-420,820
R- squared	0.21	0.53

\*\*\* indicates that the coefficient is statistically significant at 1%, \*\* at 5% and \* at 10%.

The indicator number of cows per acreage presents relatively low goodness-of-fit indicators. However, results across the different regression methods are consistent and can thus be considered reliable. Several determinants have a significant influence on the variable. In fact, some household characteristics are significant in all the regressions methods: the age of the head and the ethnic affiliation. Older farmers are less likely to have a large herd per land size, reflecting the fact that they are less market oriented. On the contrary and because of historic reasons as exposed in chapter 1, Kikuyu farmers reach higher levels of intensification.

Results reveal also that land size has a major role (in seven regressions over nine): farmers with limited land size have a higher number of cows per land size. This result is consistent with the observations derived from table 5 and shows that farmers with large land sizes have more land to allocate to dairy activities *ceteris paribus*, and thus less cows per unit of land.

The other important result deals with the availability of services: dairy co-operatives (significant in eight over nine regressions) and artificial insemination (in all the regressions). Farmers for which marketing and livestock services are available in the neighbourhood reach higher levels of intensification. In fact, dairy co-operatives provide reliable outlets for the smallholders' milk while the provision of services like artificial insemination enables the dairy enterprise to run efficiently. A similar conclusion can be drawn from the result on distances: farmers located far from urban centres have lower numbers of cows per land size. This is explained by the fact that milk sales fetch a lower price in remote areas and milk marketing is more dependent on intermediaries that may not collect milk regularly. Farmers are thus less induced to keep many cows.

The climatic variable (annual precipitation over evapo- transpiration ratio) is significant as well, reflecting the importance of feed availability on the incentive to keep cows. And farmers in tick-infected areas keep less cows since risks of contracting lethal animal diseases are higher.

Finally, in some specifications, the variables controlling for the selectivity bias are significant. For all the three estimation methods, the IMR have a significant and positive effect on the number of cows per unit of land, revealing that farmers who adopted grade cows have unobservable characteristics that induced them to adopt because they were “skilled” for dairying, either because of true characteristics (e.g. previous exposure to dairying) or more motivated. On the other hand, the polynomial terms of the probability of adoption give ambiguous results: in fact, the calculation of the overall effect shows that the effect is on average positive when using the results of the WLS estimation (the median is however negative) and negative for the other two estimations (the median is negative as well). On overall, there is thus no clear-cut conclusion concerning the possible selectivity bias.

When dealing with the second indicator, milk yield per year, the results are better in terms of the goodness of fit indicators but few variables are significant. Only one variable is significant across all the estimation methods: the climatic variable. It shows that milk yields are higher in locations where climatic conditions are better. This is consistent with the view that the major constraint in dairy farming is lack of feed (Staal 1999).

### 5.3. Analysis of the results

Two major routes toward intensification have been identified: intensification by increasing the number of cows per land and intensification by reaping high milk yields per cow. The land size plays a key role in the choice of the route: farmers with large land size have fewer animals per

unit of land. On the other hand, the variable land size does not seem to play a significant role in explaining the milk yields levels. Other factors explaining the choice of the routes are the availability of services: the availability of A.I. services is positively related with the number of cows per unit of land but not with higher milk yields. Finally, the choice between routes depends on climatic conditions: the most intensified households and those who rely on on-farm feeds for intensification benefit from higher rainfall (table 5).

Some specifications show that there is a selectivity process at play: farmers who adopt grade cows are relatively “better” at dairying, thus inducing them to adopt and enabling them to get higher milk yields. However, the variable controlling for the selectivity bias does not have a significant effect in the majority of the specifications, thus providing an incomplete conclusion.

## **6. Conclusions of chapter 5**

The analysis in this chapter seeks to answer two questions: who intensifies and how do farmers intensify their dairy activities? The major obstacle faced in the analysis is the assessment of the quantity of milk produced since milk production varies daily and thus had to be estimated in a first step using a production function approach. Results presented in section 3 are consistent with the literature and show a declining lactation curve after calving. In a second step, the yearly milk production is estimated using the results of the econometric estimation. Levels of intensification in dairy farming are then computed per farm.

The econometric results of the intensification level present low indicators of goodness of fit; nevertheless two facts emerge from the analysis. The land size plays a major role both to explain the intensification process and to understand the routes toward intensification: in fact, farmers with large land size are characterised by lower level of intensification and their route is through fewer cows per acreage but not through higher milk yields. This result suggests that high milk yields are not only achieved on large farms, but farmers with limited land sizes succeed as well in reaching high levels of milk production per cow.

Another key factor is the availability of services: artificial insemination services encourage intensification through keeping more cows (per land size). This result is particularly worth emphasising considering the declining trend in the use of these services due to the on-going privatisation of services.