Chapter 3

Development Pathways in Medium-to High-Potential Kenya: A Meso-Level Analysis of Agricultural Patterns and Determinants

Frank Place, Patti Kristjanson, Steve Staal, Russ Kruska, Tineke deWolff, Robert Zomer, and E. C. Njuguna

The highlands of East Africa have been endowed with a combination of moderate temperatures, adequate rainfall (falling in two distinct seasons for much of the highlands), and productive soils that make the region one of the best suited for agricultural development in all of Africa. As a consequence, the area has a long history of human habitation and supports some of the highest rural population densities in Africa (Hoekstra and Corbett 1995; Pender, Place, and Ehui 1999).

The good news is that in some areas in the highlands, it is clear that land use change has been part of a productive and sustainable pattern of agricultural development. Two examples of successful agricultural intensification are found in the central highlands of Kenya. The Mt. Kenya highlands have been the home for many studies (see Chapter 8) that have found improvements in living conditions and land quality. Similarly, the nearby Machakos District had a high prevalence of soil erosion, pasture degradation, and deforestation with very low agricultural productivity and income in the 1930s and was considered to be overpopulated (Tiffen, Mortimore, and Gichuki 1994). By 1990, however, the population had increased fivefold. Surprisingly, not only was there less resource degradation, but the value of agricultural output per head was estimated to be three times larger than 60 years earlier. A recent Government of Kenya study also suggests that most of the districts that comprise the central highlands of Kenya (Nyeri, Kiambu, Kirinyaga, and Meru) have much lower levels of poverty than other rural areas of Kenya (Ministry of Planning and Finance, Government of Kenya 2000). This is corroborated by a recent study by Tegemeo Institute and Michigan State University showing agricultural productivity, value of production, and incomes per hectare to be significantly higher in central Kenya than elsewhere (Argwings-Kodhek et al. 1999).

The bad news is that the trend in the majority of the highlands appears to be a downward spiral of increasing population pressure and land degradation, stagnant or declining agricultural production, and entrenched poverty (Cleaver and Schreiber 1994). In many parts of the highlands people are now living in abject poverty. Over 50 percent of the rural population in western Kenya lies below the poverty line (Ministry of Planning and Finance, Government of Kenya 2000). In many of these same areas, poverty has been accompanied by resource degradation. Reductions in woody vegetation, declines in soil fertility, and increases in soil erosion appear to be the norm in much of western Kenya, where recent studies in the Nyando River Basin have found a high proportion of physically degraded land as a result of poor land management practices (Shepherd and Walsh 2001).

The success stories are few and far between. Poverty and land degradation characterize most of the highlands. Why is this so? Nonfarm employment opportunities are growing slowly in East Africa, continuing to place pressure on agriculture to support roughly 80 percent of workers (World Bank 2002b). The road to sustainable livelihoods for rural households is difficult. The agricultural options available to communities are conditioned to some, perhaps a large, extent by physical and climatic conditions. But the policy environment plays a key role in the structure and performance of supporting systems to agriculture (e.g., extension) as well as in the promotion of markets for agricultural goods and services.

The key development challenge this chapter addresses is how the cases of successful intensification can be replicated or adapted in the wider highlands to overcome widespread poverty and land degradation in a manner that leads to sustainable improvement in livelihoods. What are successful land uses and management strategies, and are they feasible only in certain physical and climatic environments, or can they be catalyzed in diverse areas given proper market development?

In particular, we examine the following hypotheses:

- The prominence of high-value agricultural enterprises is not only predicated by climatic conditions; market development plays a significant role.

- Market development has a greater influence on higher-value agricultural enterprises in more favorable zones.

- Higher-value agricultural enterprises (cash crops, dairy, and tree growing) are associated with greater wealth.

- Higher-value agricultural enterprises (cash crops, dairy, and tree growing) are supportive of improved natural resource management.

The remainder of the chapter is as follows. Section 2 describes the methodology involved in the collection and analysis of the data. Section 3 identifies different development domains for rural Kenya based on population pressure, agricultural potential, and market access. Section 4 describes different types of land use strategies pursued by rural communities in Kenya. An analysis of the major underlying factors associated with observed land uses is found in Section 5. Section 6 evaluates the effects of different agricultural enterprises on wealth and natural resource management. Finally, Section 7 summarizes the key findings and discusses ways forward for policy.

Methodology and Data Sources

In this section, we first present the empirical models tested, followed by a description of the specific variables available. The final subsection then provides more detail on the types of statistical tests undertaken to test the hypotheses.

Developing Econometric Models

The first step is to better understand the types of agricultural strategies adopted by farming communities in the arable areas of Kenya. Our unit of analysis is the community or landscape scale, and thus we are interested in the choice of the scale of cereal production or its proportion of area versus cash crops versus woodland, for example. Decisions for land use are taken at the household level for the most part. Household decisions are determined strongly by the profitability or, more generally, provision of utility of different land uses (e.g., Chomitz and Gray 1996; Place and Otsuka 2000; Nelson and Geoghegan 2002). Farm-level profitability depends on the prices of inputs and outputs faced by the farmer and, in the case of imperfect markets, her endowment of resources, including management skills. If these concepts are projected to a larger scale, agricultural profitability depends critically on production possibilities that are in turn dependent on agro-ecological conditions. Data on prices of inputs and outputs would be ideal to include in a model. In their absence, factors hypothesized to determine prices may be market access and factor ratios (e.g., captured by population density).

One may therefore posit a model of land use at the community level as follows:

$$\text{Land use} = f(\text{conditioning factors, driving forces}) \qquad (1)$$

where conditioning factors are the exogenous physical and climatic factors such as soil type, rainfall, altitude, and temperature, and driving forces are the more dynamic drivers of landscapes such as population density and market access.

If the conditioning factors explain an overwhelming proportion of the variation, this implies traditional development and transfer of agricultural technologies appropriate to specific agro-ecological zones may well be a sufficient strategy for agricultural development. The interpretation of the coefficients for the driving forces may be made difficult if the driving forces are related to the conditioning factors. For instance, Chomitz and Gray (1996) discuss the likelihood of endogeneity of road development in the case of Belize. Thus, in order to understand the importance of individual variables, it would be appropriate to account for such relationships among the explanatory variables. As will be explained below, the availability of variables at a meso level (i.e., the middle levels roughly from village to district) is poor. This turns out to be the key constraint to ideal causal analysis because many of the key variables related to land use, such as climate, land tenure, population pressure, and market access, vary significantly at this meso scale.

If effects between conditioning factors and driving forces can be isolated, results from equation (1) will enable the confirmation of the importance of population density (through effects on factor ratios) and market access on agricultural strategies. Of particular importance will be the market variables because these are more clearly related to policy decisions. Another challenge for modeling land use is that there are several possible land use outcomes. This suggests that a system-of-equations approach is appropriate. The practical difficulty of implementing this is the lack of variables to identify the different land use equations. Indeed, as will be shown below, the major limitation of analyses covering wide geographic areas is lack of available variables.

Analysis of the land use model would provide some insights into what types of land uses are observable or attainable under various conditions and, to some extent, how such systems can be promoted. However, it does not provide evidence of why certain land uses should be promoted over others. For this, it would be important to evaluate the land use systems in regard to their impacts on productivity, poverty, or resource management variables, which are of importance to individuals and society. For instance, are cereal-dominated systems more productive than others? Are they linked to lower poverty rates, or do they lead to better-managed soils? Such an analysis is shown in equations (2a) and (2b), which use poverty as an example:

$$\text{Poverty} = f(\text{conditioning factors, driving forces}) \tag{2a}$$

$$\text{Poverty} = f(\text{conditioning factors, driving forces, land use}) \tag{2b}$$

The effects of land use on poverty (or other indicators of utility) may not be straightforward because the nonfarm sector may play a vital role in providing incomes to families. Of particular interest will be to identify the specific types of land use systems that are found to generate high levels of wealth. This question may be difficult to disentangle because the direction of causality between land use and poverty levels is ambiguous from a theoretical point of view. In other words, it is conceivable that more wealth leads to different choices as to which crops to sow or livestock to buy, but which crops or livestock products you produce may also influence how well off you are.

We present findings regarding the relationship among different land uses, wealth, and the percentage of tree cover. However, reliable meso-level data for other indicators of degradation, production, or productivity do not exist beyond small geographic areas. The regression results from equation (1) may be interpreted as identifying relationships from the explanatory variables to the land use variable because the explanatory variables are exogenous, and some of them are not modifiable at all. However, we are unable to disentangle the direction of causality between land use (and implied enterprise choice) on the one hand and poverty and environmental indicators on the other hand. In that case, we merely attempt to describe associated patterns that merit further attention.

The Data Set

The data used in the analysis reflect different spatial units and are drawn from different sources. Many of the variables, including the land use variables, are generated from aerial photos of 45-hectare-sized areas. Because all the land use data are generated from the aerial photos, this 45-hectare area is the basis for almost all the statistical analysis and is what we refer to as the "site." Others may relate to square kilometer resolution or data collected at divisional level (administrative unit below a district). These are described in more detail. In all cases, variables are geo-referenced, allowing them to be scaled up or down to link with other data. The aerial photos are geo-referenced in a crude way. The center point for the photo is identified before the flight, but the actual photo may deviate slightly from this as a result of human error in flying and taking the photo.

Available data on physical and climatic conditioning factors include altitude, slope, rainfall, length of growing season, temperature, and the precipitation to potential evapotranspiration ratio (Corbett et al. 1999). These variables are available for all of Kenya and have been catalogued by ILRI (2003). These variables constitute the most relevant variables for agriculture, with the possible exception of soil type. However, there is no reliable, recent, and high-resolution information on soil types for Kenya.

There are potentially many important driving forces behind land use decisions. At the national level, these include peace and security, economic stability, exchange rates, financial liquidity, urban growth, and facilitation of agroprocessing. Such factors would be critical in a dynamic model of land use change or intensification. However, because our land use data are from a single year, these national variables become constant across all our sites. Thus, we focus on driving forces that differ across sites within Kenya. Among these are population density and growth, market access and growth, rural nonfarm growth, presence of development projects, ethnicity, and effectiveness of extension programs. However, our data set includes population density and various measures of market access only. There is insufficient data coverage for information on the other driving forces. Population density information pertains to 1989 (the last census that has been made available to the public) and is available at the sublocation level (lowest administrative unit) for all of Kenya. Market access surfaces have been generated by ILRI and ICRAF and include distances to different types of roads and urban markets as well as estimates of travel times to different urban areas. These variables are available for the medium- to high-potential areas of Kenya.

Land use variables were generated from interpretation of low-altitude aerial photographs (1,500 meters above ground level) that were taken toward the end of the long rainy season of 1997 (May–June). Each photo represented 45 hectares, and photos followed a grid pattern at 2.5-kilometer intervals along transects that were separated by 5 kilometers. So although there is a good density of coverage, it is not complete, and we cannot produce land use surfaces. Thus, the ability to generate ex situ spatial variables, such as land use in adjacent areas, is inhibited. The original flight coverage included all areas where maize might be grown, so this excluded the north of Kenya as well as the more arid areas of southern and eastern Kenya. The available photos number over 8,200, but because of financial constraints, 5,546 photos from 30 districts were interpreted. The excluded photos are largely from the drier, lower-elevation areas. Interpretation of land use variables was done by projecting the photos onto a 100-dot grid paper on a wall and counting the number of times a certain land use intersects with a dot.[1] Tree canopy cover was calculated by visual inspection over the entire photo. Finally, data on cattle density are from 1993–97 division-level reports of the Ministry of Agriculture and Livestock Development.

The low-altitude aerial photography allowed very detailed interpretation of land use, including specific type of crop, for example, maize, sorghum, or potato. A total of 97 land use or cover variables could potentially be distinguished, and these include not only different crops but nonagricultural land covers such as water bodies, roads, and man-made structures. Because the photos covered entire 45-hectare tracts, there is considerable noncultivated area in the resulting land use data. To make our analysis tractable, we aggregated land use into a handful of discrete cases: maize, other staples, legumes, cash and horticulture,[2] grazing and pasture land (including planted forages), wooded land, other land, water, and man-made structures. In our analyses, we focus mainly on explaining differences in maize, cash/horticultural production, and woodlots, as other land types were neither common nor sizable. These calculations are made for each of the 5,546 images. We also attempt to explain the current intensity of cattle and dairy cattle raising. These data are available from the International Livestock Research Institute (ILRI) at the divisional level (a district may be composed of 5–10 divisions).

High-resolution data on poverty, productivity, and natural resource management are not yet available for wide areas of Kenya. Thus, we again turned to the aerial photographs to generate such information. For poverty, we used the proportion of roofs that were high value (tile or tin) as opposed to thatch. Although there are certain cultural preferences as to roof type, this variable is widely considered to be associated with more robust poverty measures. As for natural resource management, the percentage of land under tree canopy cover is the only useful variable at our disposal for each of the 5,546 sites. We attempted to measure soil degradation by visual identification of gullies. However, this proved to be feasible only in the drier areas because sheet and gully erosion could be hidden under vegetation in the more humid areas.

Delineation of Kenya into Development Domains

In this section, we focus on the highlands of Kenya, defined as areas exceeding 1,200 meters above sea level. Most, but not all, medium- to high-potential land is found in the highland areas. Figure 3.1 (see color insert) shows that the highlands cover much of the southwest quarter of Kenya. The easternmost area is known as the central highlands, and the westernmost area is the western highlands. The western highlands extend almost all the way to Lake Victoria, which is at 1,100 meters above sea level. In between the two is the rift valley, which, although it lies below the enveloping higher hills, still falls into our definition of the highlands. In addition to this large contiguous area, there are fragmented highland areas extending to the southeast and to the northwest.

In order to define distinctive and meaningful development domains, the highlands were partitioned into zones according to agricultural potential, population density, and market access (defined in detail below). These three variables are selected because they are expected to have a significant influence on agricultural strategies or pathways (this has been shown in other studies, including Pender, Place, and Ehui 1999; Kristjanson et al. 2002). Agricultural potential has an obvious link to agricultural strategies in that it essentially defines what options are feasible from a technical point of view.[3] Market access influences the extent to which agricultural commodities can be marketed and inputs and services obtained and is expected to influence the degree of adoption of commercial enterprises. Because population density is a general proxy for average landholding size, it is expected that high population density will lead to the adoption of more labor-intensive strategies and, according to Boserup (1965) and Ruthenberg (1985), will lead to more intensified agriculture. Population density can also be seen as reflecting market demand for local goods and services.

The actual variables used in the delineation of Kenya's highland areas were:

1. Agricultural potential: total precipitation/potential evapotranspiration ratio (or P/PE, where the numerator and denominator are both measured in millimeters).[4]

2. Market access: travel time (by vehicle in minutes) to the nearest urban area of at least 2,500 people, using assumed mean travel speeds for the main road types.

3. Population density: population density from the 1989 census.

Identifying cutoffs for each of these variables is a bit subjective. We undertake two separate exercises, one that attempted to give a more realistic yet complex picture of the variation across highland sites and another to provide a more simplistic but manageable view. The first retained four categories of population density, four levels of market access, and four agricultural potential zones and in the end found 61 different development domains in the highlands, 28 of which represented at least 1,000 square kilometers. These are not discussed in detail but serve as a reminder that the highland landscape is complex and varying. The multivariate analyses that follow later in this chapter will take this into full account, though for descriptive purposes we now simplify the picture.

The simplified delineation of development domains assumes only two categories for each of the three variables (high and low), which in combination can yield a maximum of eight distinct outcomes. The cutoffs used were 0.75 for P/PE, 30 minutes in travel time to a small center of at least 2,500 persons, and 200 persons per square kilometer. All eight possible outcomes do in fact emerge, and these are depicted in Figure 3.2 (see color insert). The different green shades are areas with relatively good market access, while those with red shades have relatively poor market access. Figure 3.2 clearly shows that the highlands near Nairobi, as well as those in the densely populated western highlands, have good access to urban markets. Market access is worse on the northernmost and southernmost reaches of the highlands. Among the low-market-access areas, almost all have low population density and low agricultural potential. There is much more variation in population density and agricultural potential within the high-market-access zones.

This is borne out by the data in Table 3.1. Of the low-market-access areas, 81.5 percent are also characterized by low agricultural potential and low population density. On the other hand, the high-market-access areas have large areas with low agricultural potential combined with low population density, high agricultural potential combined with low population density, and high agricultural potential combined with high population density. Contrasting across population density class, the high-population zones are dominated by high agricultural potential and

Table 3.1 Description and importance of development domains in the Kenya highlands

Development domains			Importance of development domains		
Agricultural potential	**Market access**	**Population density**	**Total area [km^2 (%)]**	**Total population (%)**	**Population density**
Low	Low	Low	44,599 (32.9)	734,897 (4.6)	16
Low	Low	High	309 (0.2)	102,279 (0.6)	331
Low	High	Low	31,550 (23.3)	1,736,525 (10.8)	55
Low	High	High	5,691 (4.2)	3,453,481 (21.6)	607
High	Low	Low	9,481 (7.0)	372,802 (2.3)	39
High	Low	High	359 (0.3)	126,070 (0.8)	351
High	High	Low	25,728 (19.0)	2,312,963 (14.4)	90
High	High	High	17,848 (13.2)	7,177,320 (44.8)	402
Total			135,565	16,016,337	118

high market-access characteristics (73.7 percent), followed by low agricultural potential but high market-access features (23.5 percent). Conversely, the low-population-density zones are distributed across different levels of agricultural potential and market access. Finally, if one begins with agricultural potential, it is obvious that there are no dominant patterns between areas of high and low potential.

In terms of population size, four development domains stand out:

1. high agricultural potential, high market access, high population density (7.2 million people)

2. low agricultural potential, high market access, high population density (3.5 million people)

3. high agricultural potential, high market access, low population density (2.3 million people)

4. low agricultural potential, high market access, low population density (1.7 million people)

These correspond reasonably well to the most important domains in terms of area covered, with the exception that the largest development domain in terms of area is the low-agricultural-potential, low-market-access, low-population-density domain (44,599 square kilometers), which ranks only fifth in terms of number of people covered. The domain categories are useful for descriptive purposes. However, for econometric analyses, we decompose the domain categories into the respective component variables. This is done to distinguish the effects of specific variables and to reduce the confounding influences of exogenous (e.g., rainfall) and endogenous (e.g., population density) variables.

Description of Land Use in Kenya

From this section onward, we make a slight departure from the exclusive focus on the highlands. Data on land use are available for the highlands and other areas, primarily the slightly lower-lying areas adjacent to the highland areas. We have included these additional sites (down to 1,000 meters above sea level) for several reasons: (1) many of the same agricultural enterprises are found in these outlying areas, (2) they enable the analysis to be enriched by more variation in conditioning factors and driving forces, and (3) there is no accepted definition of highlands.

Table 3.2 Percentage area under different land uses

Land use	Cases where observed (%)	Mean area (%)	Median area (%)
Grazing/fallow/pasture	94.4	45.2	43.0
Maize and intercrops	75.7	18.3	13.0
Bare/bush land	80.8	13.4	7.0
Traditional cash crops[a]	34.6	8.3	0.0
Wooded land	47.1	7.5	0.0
Man-made structures	67.6	2.5	2.0
Other staple foods	29.1	1.8	0.0
Water bodies	16.9	1.4	0.0
Horticulture[b]	6.0	0.3	0.0
Legumes	4.2	0.2	0.0

Note: n = 5,547 45-hectare units.

[a] Coffee, tea, cotton, sugar cane.

[b] Fruits and vegetables.

As noted in the methodology section, a large number of distinct land uses were identified in the aerial photo interpretation. In order to make the analysis tractable, we have combined the numerous observed land uses into 10 categories. Table 3.2 lists the 10 categories according to mean area, also indicating the median area and the percentage of nonzero observations (among the 5,546 photos). The mean area is based on a total area of 100 (recall the 100 gridpoints used to collect the information), so it can also be treated as the percentage of total area.

As can be seen, noncultivated land occupies the majority of land area. The largest single category is grazing, pasture, and fallow land, which is found in nearly all the 5,546 sites and has a mean percentage area of 45.2 percent. Bare or bush land occupies 13.4 percent of land, wooded land (woodlots, plantations, forests, woodlands) another 7.5 percent, and other noncultivated area 3.9 percent. Some of this land may well be part of a cultivation rotational practice, but at the time of the photos it was not under cultivation.

As for crops, maize and maize intercrops dominate in the areas covered. Eighteen percent of the landscape was devoted to maize, and the crop was found in nearly 76 percent of sampled sites (recall that the sites are restricted to medium- to high-potential areas). About one-fifth of sites show between 1 percent and 10 percent of land area in maize, and about 13 percent have over 40 percent of area in maize. When cultivated land is used as the denominator (and thus omitting cases where there is no cultivated land), the proportion of area under maize soars. Only 7.3 percent of sites with cultivation recorded no maize, and as many as 38.5 percent of these sites are characterized by complete dominance of cultivation by maize. Maize comprises 75 percent or more of cultivated area in 60 percent of the sites.

Traditional cash crops of coffee, tea, sugar cane, and cotton occupy around 8 percent of total land area. However, these crops are found in only about 36 percent of 5,546 cases. Figure 3.3 (see color insert) shows the geographic distribution of cash crops and provides a view of the area covered by the aerial photographs. As can be seen, there is a high concentration in central Kenya north of Nairobi (coffee and tea), in the western rift valley (tea), and in pockets of western Kenya (mainly sugar cane). Other crops are of only minor importance at the landscape scale.[5]

Data for cattle and dairy density (at divisional level) indicate that the mean number of cattle per square kilometer is 101 with a median of 72. Almost every site for which data are available reports the existence of cattle. Presence of dairy cattle is also nearly ubiquitous, though the numbers and density are significantly lower than for all breeds taken together. The average dairy cow density is 39 per square kilometer with the median being 20. There are pockets of high dairy cattle density: 17 percent of sites report dairy cow density of over 75 per square kilometer.

Planted woodlots (by farmers) were identified in 37.5 percent of sites. Most of these (32.7 percent of all sites) exhibited modest woodlot coverage, at 10 percent of land cover or less. Thus, only 4.8 percent of sites showed relatively high concentrations of woodlots (i.e., over 10 percent), and the mean across all sites was 2.1 percent of land cover (only a portion, therefore, of total wooded area). The overall tree canopy cover is expressed as the proportion of all area under tree canopy cover (Figure 3.4; see color insert). It is estimated from visual inspection of the aerial photo and thus independent of the land use assessments based on a grid sampling. For example, it considers tree density on wooded land as well as trees found on other land uses such as cultivated land. The mean tree canopy cover across all sites was 15.8 percent, and as many as 21.2 percent are estimated to have at least one-fifth of area under tree cover. At the other extreme, about 17 percent of sites have virtually no tree cover.

Factors behind Agricultural Enterprise Choice

Farming households in these areas typically have many choices confronting them as to how to best allocate their land. In a nutshell, they can plant maize or other cereal crops, beans or other legumes, a range of horticultural or cash crops, Napier grass or other animal feed crops, plant trees for fruit or fodder or soil fertility (or a combination), or dedicate some land to pastures. In this section, we analyze the factors associated with choice of agricultural enterprise. The particular enterprises examined are maize (including intercrops) area, cash and horticultural area, cattle and dairy cattle density, and woodlot area. Other specific agricultural enterprises are not examined mainly because of lack of prominence.

Because of a restricted number of available explanatory variables, we do not develop causal models but rather models of association or prediction (though there is little reason to believe that causality runs from land use to the explanatory variables). Essentially, although it is clear that population density and travel time to markets are related to climatic conditions (and each other), it was not possible to distinguish these "subrelationships." To compensate, a number of models were run, with and without individual variables, in order to better understand the relationships among explanatory variables and the resulting direct and indirect effects that they may have on agricultural enterprise choice. In some cases where there was high correlation among variables, one or more were removed to avoid multicollinearity problems. Because of the limited number of variables, we are also obliged to run single-equation models. This type of situation lends itself to a seemingly unrelated regression model, but that procedure offers no improvement to single-equation models when the explanatory variables are the same across equations. Further, systems equation models are more complicated when limited dependent variables are concerned.

Tables 3.3–3.6 show regression results for maize, cash crops, cattle, and woodlots, respectively. The models presented are those that include both conditioning factors and driving forces. Results from other specifications will be discussed, but are not presented in tables because of space limitations. We ran models with and without district dummy variables. Because most of the results are the same in terms of sign and significance level, we present only those from the model in which they

Table 3.3 Tobit regression of maize area and percentage of cultivated area under maize

	Maize area		Cultivated area under maize (%)	
Variable	Coefficient estimate	Significance level	Coefficient estimate	Significance level
Constant	–37.47938	0.010	0.41234	0.400
Altitude (meters)	0.014152	0.404	0.0001	0.786
Altitude squared	–0.000006	0.158	–0.00000004	0.696
Precipitation (millimeters)/evapotranspiration (millimeters) ratio (PPE)	106.0421	0.009	0.34531	0.619
PPE squared	–67.2070	0.005	–0.36486	0.368
Travel time to urban area (hours)	0.03581	0.054	0.000715	0.082
Travel time squared	–0.000057	0.007	–0.000001	0.007
Population density (persons/km^2)	0.066772	0.000	0.000025	0.894
Population density squared	–0.000042	0.000	–0.00000002	0.845
Number of observations	5,515		4,501	

Note: District dummy variables included but not reported in the table.

Table 3.4 Tobit regressions of cash crop area and percentage of cultivated area under cash crops

	Cash crop area		Cultivated area under cash crops (%)	
Variable	**Coefficient estimate**	**Significance level**	**Coefficient estimate**	**Significance level**
Constant	–94.9848	0.002	–1.60977	0.000
Altitude (meters)	0.03528	0.253	0.001169	0.055
Altitude squared	–0.00001	0.199	–0.0000003	0.077
Precipitation (millimeters)/evapotranspiration (millimeters) ratio (PPE)	163.6994	0.004	1.28654	0.235
PPE squared	–77.1062	0.017	–0.42426	0.475
Travel time to urban area (hours)	–0.07860	0.016	–0.001422	0.015
Travel time squared	0.000056	0.091	0.0000016	0.014
Population density (persons per km^2)	0.054827	0.006	0.000246	0.423
Population density squared	–0.000034	0.008	–0.00000015	0.441
Number of observations	5,515		4,501	

Note: District dummy variables included but not reported in the table.

are included. For the sake of space, the coefficient estimates on the district dummies (31 of them) are not included in the tables. In all regressions, many of the district variables turn out to be very important and significantly raise the explanatory power of the models. We also tried to account for the nonindependence of observations (i.e., spatial autocorrelation) through a clustering technique offered in

Table 3.5 OLS regressions of density of cattle and dairy cattle

	Cattle		Dairy cattle	
Variable	**Coefficient estimate**	**Significance level**	**Coefficient estimate**	**Significance level**
Constant	97.5708	0.044	–79.23615	0.196
Altitude (meters)	–0.01653	0.743	0.09026	0.006
Altitude squared	–0.0000027	0.838	–0.00002	0.018
Precipitation (millimeters)/evapotranspiration (millimeters) ratio (PPE)	112.2307	0.337	157.8382	0.096
PPE squared	–47.0445	0.513	–94.8817	0.083
Travel time to urban area (hours)	0.005150	0.960	–0.035043	0.338
Travel time squared	–0.000032	0.738	0.000026	0.475
Population density (persons per km^2)	0.09932	0.006	0.03498	0.102
Population density squared	–0.000025	0.298	–0.000015	0.224
R^2	0.403		0.549	
Number of observations	4,766		4,766	

Note: District dummy variables included but not reported in the table.

Table 3.6 Tobit regressions of area under woodlots

	Model 1		Model 2	
Variable	Coefficient estimate	Significance level	Coefficient estimate	Significance level
Constant	−37.02432	0.000	−36.28394	0.000
Altitude (meters)	0.015225	0.010	0.012687	0.029
Altitude squared	−0.0000036	0.020	−0.000002	0.123
Precipitation (millimeters)/evapotranspiration (millimeters) ratio (PPE)	35.34481	0.008	34.2274	0.009
PPE squared	−16.68914	0.014	−15.4829	0.020
Travel time to urban area (hours)	−0.010756	0.083	−0.00621	0.287
Travel time squared	0.000006	0.346	0.000002	0.742
Population density (persons per km^2)	0.019333	0.000	0.01654	0.000
Population density squared	−0.000011	0.000	−0.000009	0.000
Slope (average change in meters)	0.036862	0.351	0.067665	0.091
Area under off-farm woody vegetation (percentage of 45-hectare site)			−0.17840	0.000
Number of observations	5,515		5,515	

Note: District dummy variables included but not reported in the table.

STATA. This enabled us to treat observations from the same district as non-independent and thereby generate more conservative (i.e., higher) estimates of standard errors for the coefficient estimates.[6]

The dependent variables have been discussed previously, so here we briefly describe the main explanatory variables included in the models. Altitude and the precipitation-to–potential evapotranspiration ratio (P/PE) are used as the conditioning factors. Other variables such as rainfall and length of growing period are essentially captured by P/PE. The squares of each are also used because of the expected reduction in suitability of many crops at very high altitudes or very humid conditions. The mean altitude across all observations is 1,615 meters above sea level. About 22 percent of cases are below 1,200 meters and almost 38 percent are above 1,800 meters. The mean P/PE value is 0.75, which is relatively favorable for a range of crop growth. Population density and travel time to the nearest urban center of at least 2,500 people (and their squared terms) are the two driving forces included. The mean human population density (in 1989) is 233 per square kilometer. Fewer than 18 percent of the sites had relatively low population densities of below 50, and 27 percent of sites had densities of over 250. The mean travel time to the nearest town of 2,500 persons is 1.85 hours. Just about 14 percent of the sites had to travel for 3 hours or more to reach a substantial market, whereas 12 percent could reach one in approximately one-half hour. A measure of slope is also used in the woodlot estimation, which measures the difference in elevation at the site compared to adjacent areas.

Tables 3.3 and 3.4 show the results of regressions on maize and cash crop area as well as the percentage of cropped area under each. We ran models with the absolute land area and the percentage area to better understand the influence of population pressure. For instance, it is expected that population pressure will lead to greater cultivated area, possibly of all types of crops. But the absolute area equations are able to assess the population densities at which this expansion diminishes and possibly stops. On the other hand, the percentage share regressions focus more on the relative importance of cash crops and maize as population density increases. Both types of information are important. Because there are zero values, a high number in the case of cash crop area, censored (Tobit) regressions are used.

In terms of total area, the results show that the maize and cash crop area increases with improved climate and greater population density, but at diminishing rates for both. So favorable climate and population pressure induce conversion of nonagricultural land to maize and other agricultural enterprises. Market access has different influences on maize and cash crops with improved access associated with higher cash crop area and percentage area under cash crops.[7] The opposite effect occurs for maize.

Some interesting results arise from the models in which maize and cash crop area as a percentage of all cultivated area are the dependent variables. The effect of P/PE becomes insignificant, and for cash crops, altitude becomes important. Thus, it is particularly at the relatively high altitudes where cash crops displace maize systems. But the average rainfall has little effect on the ratio of cash to maize crop production. Interestingly, although population pressure leads to expansion of cropped area, it does not directly influence the balance between maize and cash crop cultivation. Such a result has been found in a number of household studies that indicate dual pressures for food production and income generation are equally felt by households constrained by small farm size (e.g., Owuor 1999).

Separate regressions included interaction terms for P/PE and travel time to market variables in order to test whether market access has a greater impact on cash crop area in areas with favorable climates. Indeed, this was confirmed: the negative and significant coefficient estimate indicates that the effect of market access is greater where climate is more favorable.

Table 3.5 shows the factors influencing overall cattle density as well as the density of dairy cattle in particular. Cattle density is not highly linked to the included variables. This is because of the different types of cattle breeds and production systems in Kenya. Some less favorable zones for cropping are attractive habitats for local zebu. Similarly to cash crops, climate and altitude play a role in predicting dairy cattle density. Surprisingly, market access and population density did not have added effects on dairy cattle numbers. This contrasts to a previous wide-scale household-level study of dairy production in Kenya (Staal et al. 2002).[8]

Table 3.6 shows results from regressions aimed at explaining factors influencing a household's decision on how much land to devote to woodlots. Model 2 differs from Model 1 by the inclusion of off-farm wooded area; the hypothesis is that such areas may reduce incentives for farmers to plant their own woodlots. Woodlots are promoted by favorable climate, population density, and to a lesser extent by market access. This is not surprising because, in Kenya, woodlots provide important sources of income, particularly fast-growing eucalyptus poles. Further, there was only weak support for the hypothesis that farmers plant trees on more sloped land because of the comparative advantage of trees over crops on such land. Finally, it does appear that farmers have access to trees off-farm and that this access greatly reduces incentives for investing in woodlots on their farms.

Impacts of Agricultural Enterprise Choice

A natural reaction to the analysis above may be to ask why any particular agricultural enterprise might be preferred over another. In other words, is there any evidence that certain agricultural enterprises are more productive, profitable, or take better care of the natural resource base than others? The one proxy variable calculated from the aerial photos relevant to profits or poverty was the percentage of roofs made of high-quality material (i.e., tin or tiles). This variable is often used as one component of a wealth index of households. Table 3.7 shows the results of censored regressions to explain this ratio. Model 1 contains the same explanatory variables as used in the agricultural enterprise regressions. Model 2 adds three enterprise variables: the percentage of area under cash crops, the density of dairy cattle, and the area under woodlots. These are endogenous variables, but the intention here is to identify whether these choice variables appear to have an impact on poverty and the environment and therefore draw attention for further investigation.[9]

This wealth indicator is related in much the same way to the conditioning factors and driving forces as were the higher-value agricultural enterprises. Agricultural potential, market access, and population density all have a significant influence on household wealth (and the expected signs). Only altitude and travel time remain significant after the three enterprise variables are added in. Cash crops, dairy cattle, and woodlots are each positively related to high-quality roofs, though cash crops are only weakly significant. Although it is not possible to state unequivocally that these land uses promote wealth accumulation, this finding strongly suggests that such enterprises are important ingredients in wealth-generating processes.

Table 3.7 Tobit regressions of ratio of high-quality roofs to total roofs

	Model 1		Model 2	
Variable	Coefficient estimate	Significance level	Coefficient estimate	Significance level
Constant	−0.69617	0.000	−0.48198	0.001
Altitude (meters)	0.000897	0.000	0.000821	0.000
Altitude squared	−0.0000002	0.004	−0.0000002	0.004
Precipitation (millimeters)/evapotranspiration (millimeters) ratio (PPE)	0.65680	0.046	0.29138	0.341
PPE squared	−0.36297	0.052	−0.19265	0.254
Travel time to urban area (hours)	−0.000783	0.002	−0.00077	0.006
Travel time squared	0.0000007	0.088	0.0000005	0.308
Population density (persons per km^2)	0.000216	0.018	0.0001198	0.158
Population density squared	−0.0000001	0.137	−0.00000007	0.262
Density of dairy cattle (number per km^2)	0.000347	0.037		
Percentage cropped area under cash crops	0.001168	0.141		
Area under woodlots (percentage of 45-hectare site)	0.00850	0.001		
Number of observations	4,181		3,640	

Note: District dummy variables included but not reported in the table.

A study by Rommelse (2001a) in western Kenya supports the importance of these enterprises. Of the top ten most common income-generating agricultural enterprises, four were livestock related (e.g., eggs and chickens), three were tree related (e.g., fruits and poles), and the top category was a horticultural enterprise (vegetables). In addition, Nicholson et al. (1999) report significantly higher incomes among dairy producer households in coastal Kenya, compared to nondairy households.

Many of these results are supported or strengthened by recent nationwide research by Tegemeo Institute (Egerton University) and Michigan State University. First, Nyoro (1999) presents data from the Ministry of Agriculture and Rural Development that shows recent (1990–95) expansion of cropped area to be fastest for horticultural crops, followed by traditional cash crops and finally by maize. Second, Owuor (1999) found that commercialization (share of total production that is marketed) is strongly related to value of crop per hectare. This explains a high proportion of the differences in per-hectare value of production both inter-regionally as well as between households within regions. Last, work by Argwings-Kodhek et al. (1999) demonstrate the large differences in income across different farming zones in Kenya. In areas with high concentrations of traditional cash crops or horticultural crops, such as the central highlands, households annually earn about $1,780 per farm from crops and livestock. In the western highlands, where farms are more or less the same size (i.e., very small, less than 1.5 hectares), the average farm earnings were only $613.

A final analysis looked at the impact of land use on the percentage of tree cover across the entire landscape of each site (i.e., the entire 45-hectare photograph). As shown in Table 3.8, tree cover is highest in the lower altitudes and then decreases, but at slower rates, as altitude increases. Tree cover is also much greater in areas with more sloping land, perhaps because of the difficulty in cultivating these lands. Climate, controlled for altitude, was not related to overall tree cover. Market access and population pressure were negatively related to tree cover, as would be expected. Model 2 tests for the relationship between high-value agricultural enterprises and tree cover. It can be seen that all three variables (dairy cattle density, woodlot area, and percentage of cultivated land under cash crops) are positively associated with tree cover, and in the case of woodlots, the coefficient is significant. This means that these agricultural enterprises, which positively impact on wealth, have a neutral or positive impact on vegetation cover as well. Whether this is primarily because of effects within agricultural land or due to pressures on resources outside of agricultural land is not clear, however.

Table 3.8 Tobit regressions of percentage tree cover

	Model 1		**Model 2**	
Variable	**Coefficient estimate**	**Significance level**	**Coefficient estimate**	**Significance level**
Constant	26.695	0.039	41.0265	0.000
Altitude (meters)	–0.06675	0.001	–0.05232	0.003
Altitude squared	0.000023	0.000	0.000015	0.003
Precipitation (millimeters)/evapotranspiration (millimeters) ratio (PPE)	8.15459	0.713	–1.94683	0.919
PPE squared	8.78486	0.465	8.45311	0.419
Travel time to urban area (hours)	0.067249	0.001	0.025639	0.008
Travel time squared	0.000066	0.000	0.000008	0.328
Population density (persons per km^2)	–0.02023	0.053	0.011158	0.061
Population density squared	0.000014	0.056	0.0000069	0.131
Slope (average change in meters)	0.92345	0.000	0.56921	0.000
Density of dairy cattle (number per km^2)			0.016944	0.3194
Area under woodlots (percentage of 45-hectare site)			0.45054	0.000
Percentage of cultivated area under cash crops			10.61746	0.369
Number of observations	5,515		5,515	

Note: District dummy variables included but not reported in the table.

Conclusions

Major Empirical Findings

The major empirical findings can be summarized as follows:

- As expected, climate and altitude are important in explaining land use, but other factors also play important roles.
- Population pressure positively influences the area under cultivation but does not automatically lead to adoption of higher-value crops.
- Good market access is critical for promoting production of higher-value agricultural enterprises, especially in the more favorable climate zones.
- Dairy and woodlots contribute to wealth generation (as measured by house quality) and at the same time have neutral or positive effects on overall tree cover.

Methodological Challenges

Our analysis focused on the visible side of rural livelihoods, namely agriculture. However, it is well known (e.g., Argwings-Kodhek et al. 1999) that the nonfarm economy plays a critical role in household strategies directly as well as indirectly through agriculture. This large sector could not be addressed by this analysis. A second limitation is the use of single-equation models that ultimately may show only patterns of association rather than causal relationships. Further progress in this area is constrained mainly by lack of breadth in variables that exist for such wide coverages at sufficient levels of disaggregation. Nonetheless, many of the relationships demonstrated in this analysis are supported by other studies. Qualitative research, such as focus group discussions, have proven very valuable in disentangling timelines of change in variables, which ultimately may enable the distinguishing of causes from effects (e.g., Kanbur 2003; Krishna et al. 2004).

A third limitation concerns the existence of spatial autocorrelation in our dataset without sufficient treatment in our statistical analysis. In case of spatial autocorrelation, the information content of the sample is lowered, rendering it less efficient than uncorrelated counterparts, so parameter estimates are inefficient, although asymptotically unbiased (Anselin 2001). Moreover, the omission of a spatially correlated and important variable may result in biased estimates. We have partially dealt with this issue through clustered regression techniques. Further work to address spatial autocorrelation in limited-dependent variable models is required.

Policy Implications

The promotion of markets through investment in roads and other infrastructure is an obvious implication of our results. This is especially true in the more favorable climatic zones. Support for this result in Kenya for the dairy sector is seen in Staal et al. (2002), and regarding impact on household incomes, in Argwings-Kodhek et al. (1999). This broad-based intervention is a good strategy because evidence shows that farmers like to diversify among many agricultural enterprises, including food, feed, and cash crops. Having said that, there is still scope for promoting markets for long-standing and new cash crops and for disseminating information about their management. In the less favorable areas, there is the additional need to identify and develop higher-value enterprises suitable to these areas (in addition to cattle raising, which is already practiced by households) because road development does not seem to have the same strong impact with the currently available cash crops as it does in the higher-potential zones. Finally, given our results regarding the positive influence cash crops, dairy cattle, and woodlots have on wealth, the predominant role of maize in smallholder agriculture should be seriously addressed within Kenya's Poverty Reduction and Rural Development strategies, and support to these other options pursued.

Notes

1. There was an attempt to "train" GIS software to distinguish among different land uses, but this would have been enormously expensive and risky if the intention was to keep as many as 90 different land use categories. The analysts who did the job were the same who had been doing similar interpretation for over 4 years.

2. Cash and horticultural crops include industrial crops such as sugar, tea, coffee, and pyrethrum, vegetables, and fruits.

3. This is particularly the case in Kenya, where area under irrigation is minuscule.

4. An increase in P/PE of 0.1, holding temperature constant, is approximately equivalent to an increase in rainfall of 143 millimeters annually.

5. These figures match fairly well with other available farm-level surveys, except for Napier grass, which has been found to be quite prominent in many districts yet almost absent in the aerial photo interpretation (Staal et al. 1997).

6. These were the svyreg and svyintreg commands in STATA.

7. The results from the regressions show a curvilinear relationship between market access and land use. For all but a handful of observed values for market access, the effect of improved market access is indeed positive.

8. Indeed, under the assumption of independent observations, market access becomes highly significant in our estimations.

9. As noted earlier, ideally one would use a two-stage procedure to remove biases that may emerge. However, lack of available exogenous variables at this scale prevent such an analysis.