This is to certify that the
thesis entitled

TRIPLE HURDLE MODEL OF SMALLHOLDER PRODUCTION AND
MARKET PARTICIPATION IN KENYA'S DAIRY SECTOR

presented by

William J. Burke

has been accepted towards fulfillment
of the requirements for the

Master of Science degree in Agricultural Economics

Major Professor's Signature

March 25, 2009

Date

PLACE IN RETURN BOX to remove this checkout from your record.
TO AVOID FINES return on or before date due.
MAY BE RECALLED with earlier due date if requested.

DATE DUE	DATE DUE	DATE DUE

TRIPLE HURDLE MODEL OF SMALLHOLDER PRODUCTION AND MARKET PARTICIPATION IN KENYA'S DAIRY SECTOR

By

William J. Burke

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

MASTER OF SCIENCE

Agricultural Economics

2009

ABSTRACT

TRIPLE HURDLE MODEL OF SMALLHOLDER PRODUCTION AND MARKET PARTICIPATION IN KENYA'S DAIRY SECTOR

By

William J. Burke

In Kenya, strong demand and the fact that most of the nation's 3 million dairy cattle are in the hands of smallholders provides a tremendous opportunity for households to participate in the dairy market and increase rural incomes. Unfortunately, recent output has not kept pace with increasing demand, suggesting that barriers prevent rural farmers from tapping dairy's underexploited potential. Using 11-year panel data from 1275 smallholders, this study develops a model to determine the factors enabling smallholder participation in Kenya's dairy market, and uses the findings to identify strategies to improve dairy productivity and promote successful smallholder commercialization. Traditional double-hurdle market participation models are not adequate for addressing these objectives, primarily because they require the implicit assumption that all farmers are producers, whereas roughly 1/3 of rural Kenyan households do not produce milk in a given year. This study thus develops a "triple-hurdle" model, which allows for both non-producers and autarkic producers. Results suggest a bi-modal policy response to enable producers as well as the formal and informal purchasing enterprises to which they sell. Technical education, improved technologies, electrification, and access to credit are important to provide an enabling environment for producers. Along with the recent initiative to revive the parastatal dairy purchaser, evidence indicates that a more stable policy environment for small-scale traders, whose current market behavior is technically illegal and unpredictably regulated, would promote significant farmer response.

For Cassie and Liam

ACKNOWLEDGMENTS

I am extremely grateful to my major professor and advisor, Dr. Thom Jayne, for the amazing opportunity to work on the Kenya project, and for his guidance, insight, and encouragement throughout my work on this and many other projects. Without his support and attentive efforts this paper would not have been possible. I am also grateful to my committee members Dr. Robert Myers and Dr. Lisa Cook, whose efforts and suggestions have greatly improved my experience and the quality of this work. I would like to thank Margaret Beaver for her support and training. This study would not have been possible without funding from USAID and the TAPRA project. My thanks go to the Tegemeo institute for use of the data and to its staff for the dedication to excellence in collecting survey data.

I would like to thank all of my fellow students at MSU, especially Joshua Ariga and Joleen Hadrich for their comments and advice, as well as Elliot Mghenyi for taking a considerable amount of time to introduce me to the data used in this study.

Since childhood my parents have provided me with the greatest loving care, direction, and opportunity that any person could hope for. You convinced us that we really could be anything, so thank you. I only hope that I can be anywhere near as good a parent. Thanks to my brothers and sister for unending support. Thanks to Marty for showing me it can be done and encouraging me along the way.

Thanks to Liam for the extraordinary motivation, and thank you most of all to my best friend and wife, Cassie, for being with me through it all. I will always admire you, and you amaze me almost every day.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

KEY TO SYMBOLS OR ABBREVIATIONS

AEZ	Agro-Ecological Zone
APE	Average Partial Effect
hh	household
IMR	Inverse Mills Ratio
IPW	Inverse Probability Weight
KCC	Kenya Creameries Company
kg	kilogram
km	kilometer
Ksh	Kenyan Schilling
MLE	Maximum Likelihood Estimator
mm	millimeter
MP	Market Participation
TAMPA	Tegemeo Agricultural Monitoring and Policy Analysis
USAID	United States Agency for International Development

1. INTRODUCTION

Structural transformation and economic growth have regrettably eluded much of Africa. In theory, increased productivity on the farm will lead to lower food prices, raise the disposable incomes of food consumers, make labor available for a growing industrial sector, and initiate the structural transformation processes in a self-perpetuating cycle of growth (Johnston and Mellor, 1961; Mellor, 1998).[1] Evidence of this cycle was well documented through Asia's Green Revolution, but the pattern has yet to emerge in Africa. Although this theory is widely accepted, specific determinants of smallholders' ability to participate and succeed in markets with the most growth potential are often mysterious.

In many parts of Africa the dairy sector has been identified for its potential to increase the income generating productivity of smallholders' assets (Walshe et. al., 1991; Staal et. al. 1997; Kodhek and Karin, 1999; Thorpe et. al. 2000). In Kenya, for example, preferences for local "raw" milk, and the fact that most of the nation's 3 million dairy cattle (85% of East Africa's dairy cattle population) are in the hands of smallholders, provide a tremendous opportunity for households to participate in that growth market (Staal and Mullins, 1996; Thorpe et. al. 2000).[2]

Indeed, roughly 69% of households in this study's sample are producing dairy in any given year. Nevertheless, in many cases dairy production does not lead to increased disposable income, as one might expect, with only 71% of producers being net sellers,

---

[1] This paradigm was subsequently adapted to encompass the direct effects of increased productivity on rural poverty as well as linkages from the farm to the rural non-farm sector (Hazell and Haggblade, 1993).
[2] "Growth markets" refer to markets for goods which require substantial initial investment (such as in education or capital), but provide higher returns than staple commodity markets, like maize. Many horticultural markets, in addition to dairy, are examples of "growth markets"

while 12% of producers withdraw from the market, and 17% of producers purchase more milk than they sell. Recently local media outlets suggest that the prevailing opinion is that the dairy sector's performance has not lived up to it's potential.[3] This underscores the importance of understanding what motivates households to produce dairy and participate in dairy markets, and why some households appear able to exploit seemingly attractive production and marketing opportunities while others cannot. Existing market participation (MP) models, however, are somewhat inadequate, since each requires that all households be producers (recall that nearly a third of the Kenyan sample does not produce dairy).

So, the objectives of this study are two-fold: (1) to develop and estimate a model eplaining the factors enabling smallholder participation in Kenya's dairy sector, accounting for non-producers, and (2) to use the findings of the model to identify strategies to improve smallholder dairy productivity and promote successful commercialization of dairy production. These objectives will be met by addressing the following research questions:

- What are the determinants of rural smallholder's ability and willingness to produce in the relatively high-value dairy sector?

- What are the determinants of a producer's role in the market as either buyers, sellers, or autarkic?

[3] See referenced material on-line at allafrica.com and africanews.com

- How do these determinants affect the level of participation, or amounts bought and sold, among participants?

For example, prior to 1992 the only legal marketing channel for dairy in Kenya was the Kenyan Creameries Company (KCC), which is effectively a parastatal firm. Since 1992, aside from household-to-household transactions, a variety of channels have become available to dairy farmers, where the KCC now shares the marketing of dairy with farmer owned cooperatives, private processors in the formal sector, and informal private traders. It is important to note that the latter group, sometimes called "hawkers," is still operating illegally. Although generally tolerated, these entrepreneurs conduct their business in an extremely uncertain policy environment (Staal et. al., 1997; Owango et. al., 1998; Kijima et. al. 2006). In response to lower than desired production nationally, the recent policy thrust has focused on the revival of the KCC (which has been in a weakened state since the dissolution of the executive board in 1998), yet there has been little empirical analysis as to how farmers respond to these various marketing channels (i.e. is this the most efficient use of government resources?).

The research questions will be addressed using an econometric model of the market participation decision. Market participation models have previously described decisions as occurring in two steps: 1) whether to participate in the market (buy or sell, versus remain autarkic), and 2) what volume will be bought or sold (Goetz, 1992; Key et. al., 2000; Holloway et. al., 2001; Bellemare and Barrett, 2006; and others). Again, however, these have limitations for the current purpose because they rely on the assumption that all households in the population are producers. While this assumption may be appropriate when applied to, say, staple grains, it is unreasonable for commodities which are not produced by a large fraction of households. Most agricultural commodities produced by smallholder farmers in fact belong in this latter category, such as horticultural products, industrial cash crops, and animal products, including dairy products in Kenya. Therefore, findings from standard MP studies understate the effects of a given determinant, because they cannot account for the impact on likelihood of production, the necessary precursor to any market related decisions. Also, in subsuming the effects of production determinants, the estimates of market participation models tend to commingle constraints on production and constraints on participating in markets, and therefore overemphasize the role of market access in explaining non-participation in markets as opposed to asset constraints or low productivity in the use of factors of production. A related issue is that many previous studies address market participation decisions as though they are made entirely *a priori*. One could argue, however, that the more appropriate discussion would be about market participation that is in part stochastic, and in some cases even unintended, determined by realized consumption and production shocks, which are in turn at least partially determined by factors exogenous to the household (e.g. rainfall or health related shocks).

This study further develops the market participation class of models to provide a three-tiered decision framework that enables a nationally representative sample to be maintained even in markets where a sub-population do not producer. The expanded framework also addresses the possibility that market participation is partially determined by exogenous shocks originating outside the household, after production decisions are made.

2. BACKGROUND OF KENYA AS A CASE STUDY

In many parts of Africa, the dairy sector has long been identified as holding potential for smallholders to increase the income generating productivity of their assets. This is particularly true in Kenya, where demand for local "raw" milk is strong domestically, as is export demand in the formal sector (Walshe et. al., 1991; Staal et. al. 1997; Kodhek and Karin, 1999; Thorpe et. al. 2000). In fact, Karanja (2003) predicted that by the end of 2008 domestic demand for dairy, driven by a growing, wealthier urban population, would outpace domestic production. While production among smallholders has increased over the past several years, local media indicates response to rising prices have been lower than expected, leading some to conclude that Kenya will be a net importer of dairy in the near future. [4]

In the World Bank's 2008 "Kenya Agricultural Policy Review," the policy debate concerning the informal dairy market is summarized as a balance that weighs increased regulation and the safety of consumers against producers' ability to thrive and provide increased productivity and output. The same report, however, states that consumers are safe under current regulations, citing that bacteria counts in the relatively unregulated "raw" milk are comparable to those in processed milk.

Raw milk, which is actually boiled prior to consumption to kill milk-bourn illnesses, has a smaller marketing margin than processed milk, leading to higher prices for the farmer and lower prices for the consumer. That is, the cost of boiling is so small relative to formal processing, that even if boiling cost is paid by the consumer both they and

---

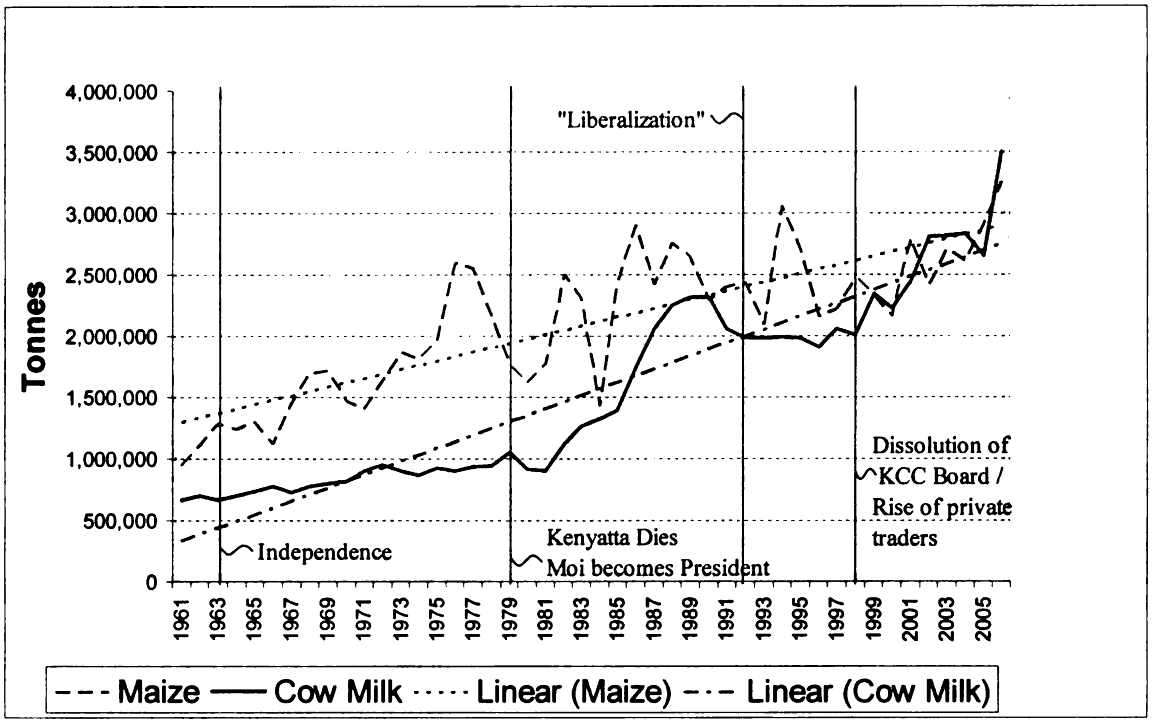
[4] See referenced material on-line at allafrica.com and africanews.com

producers receive preferable prices in the raw milk market (Staal and Mullins, 1996; Thorpe et. al. 2000, World Bank, 2008). Indeed, evidence indicates participation in the dairy market has not only been prevalent among Kenya's rural smallholders over the last decade, but that this activity is generally associated with increasing asset-wealth over time (Burke et. al. 2007). Thus, if more regulation is thought to have only marginal benefit to consumers, the remaining question is: what would provide a more attractive environment for producers?

Regarding the formal market, recent policy has focused on the revival of the KCC (allafrica.com, 2008; africanews.com, 2007). This raises an interesting point. Consider Figure 1, which uses national level data and super imposes dairy (and maize, the countries primary crop) production over a political timeline. This shows that, after an initial lull during the period where a strong KCC and growing private sector coexisted, annual production has increased dramatically since the KCC central board was dissolved in 1998 amidst corruption scandal. This calls to question whether the current policy focus is the most efficient use of government resources. Figure 1 also shows the gap in production trends between milk and maize has been narrowing steadily for decades.

That said, from the producer's perspective, as with all entrepreneurial endeavors, there are obstacles which must be scaled by small farmers to realize the potential benefits of dairy production. First, milk production, like that of any other agricultural product, is partially a stochastic process, where factors outside the farmer's control (i.e. rainfall) can affect output.

Figure 1: Production (1000 tonnes) Over Time for Milk and Maize in Kenya



Source: FAOStat web site, FAOStat database: www.faostat.fao.org

Secondly, there is substantial overhead investment required to own, house and feed dairy cattle, as well as having labor available to tend the herd. For example, high-yielding "grade" cows are considered essential to sustainable dairy farming (World Bank, 2008). In 2007 the median price of a single improved breed cow was 30,000 Ksh, which at 24% of the median annual net income of rural households is a sizable investment. Moreover, a number of veterinary services are available, such as artificial insemination, which could make an endeavor more likely to succeed, but which may also present additional barriers to entry. Once a household does enter the dairy sector, there is no guarantee of success. Milk, being highly perishable, requires either cold storage or fairly immediate market access. Also, there is a variety of hidden action problems involved in any market exchange, since the quality (milk can be diluted) and hygiene (poorly handled milk is unsafe) can vary greatly, and are not easily verifiable.[5]

In short, there are three broad classes of variables that may explain heterogeneous household response and outcomes over time to investment in dairy: (1) a stochastic production process; (2) location-specific differences in the enabling environment, including access to and performance of localized input and output markets, and agro-ecology; and (3) household specific characteristics such as knowledge and training, asset levels, etc. An important point emerging from this discussion is that successful participation in dairy marketing depends on far more than the willingness of the farmer to enter, or even his/her specific market-related conditions.

Although prior research has acknowledged the market's potential and examined the industry using regional or national data, there has been little household-level panel analysis of dairy in Kenya. Presumably, this is due to the dearth of panel data required to properly investigate markets from the household perspective. It is important, however, to understand how the household unit fits into the expanding market. The lack of empirical research identifying determinants of whether and how households are participating in Kenya's expanding dairy market leaves an important void to be filled.

[5] Despite these information asymmetries, some empirical evidence shows raw milk to be comparable in safety to processed milk, possibly because of the boiling process, or the importance of reputation building through repeated transactions and/or cooperative organizations (World Bank, 2008).

3. DATA

This study uses panel data from four surveys implemented by the Tegemeo Institute of Egerton University in Nairobi, Kenya. In 1997, the sampling frame was designed in consultation with the Central Bureau of Statistics, and contained 1,500 households randomly chosen to represent eight different agricultural-ecological zones (AEZ), reflecting population distribution. Of the original sample, 1,428 households (95%) were re-interviewed in 2000, 1,324 (88%) were re-interviewed in 2004, and 1,275 (85%) were re-interviewed in 2007. Holding consistently at or below 7% of the original sample per survey, this rate of attrition is reasonably low compared to similar surveys in developing countries, which can typically range as high as 20% (Alderman et. al., 2001).

Attrition bias should always be a concern when working with panel data, but preliminary analysis has shown the attrition in this sample does not appear to be highly systematic, and correction measures have not altered results meaningfully. Burke et. al. (2007), for example, used inverse probability weights (IPW) based on household and community characteristics. This involves estimating an auxiliary model for attrition and weighted observations with by their re-interview probability. Ultimately, the majority of re-interview coefficients were not statistically significant. Moreover, results from that study's weighted probit analysis had no more explanatory power than the un-weighted, and most results were identical to the thousandth decimal.

This nationally representative panel data gives this study the unique opportunity to examine the household with respect to dairy market participation in order to address the research questions of concern.

4. CONCEPTUAL FRAMEWORK

4.1 Market Participation

Economists have generally treated the household's decision to participate in markets as a two-step process: first, producing households decide whether to participate (buying or selling) or remain autarkic, then, conditional on participation, how much to buy or sell (Goetz, 1992; Key et. al., 2000; Holloway et. al.,2001; Bellmare and Barrett, 2006; others.). However, when considering a market such as dairy in Kenya, it is important to first acknowledge that not all households will be producers. For example, at the outset of the 10 year panel used for this study, only about 2/3 of the households were milk producers. This makes it important to add a third stage of analysis to the traditional 2-stage MP model that identifies factors influencing a household's decision whether or not to produce.

Moreover, existing models treat participation entirely as an ex-ante decision made by the household. By adding production decisions in a separate stage, however, this study will allow factors affecting production decisions to differ from those affecting market participation, thus acknowledging the fact that marketing decisions may be, in part, not an ex ante choice but often an artifact of constraints on production and/or a response to stochastic production shocks. Production decisions can only be made based on the farmer's expectations. This distinction allows the farmer's information to be updated after deciding whether to produce, but before deciding to participate (or not) in the market.

There is a drawback to imposing the assumption that production and marketing are discrete sequential decisions when using annually aggregated household data. Unlike annual crop production (where output is realized at a specific harvest time), dairy production is a temporally continuous process where annual yields are realized gradually. Imposing a sequential decision model implicitly assumes that when production decisions are made they are maintained throughout the year. Concerns over this implicit assumption can be partially assuaged by the high degree of seasonality in dairy production (which is highest during the rainy season). Of course, the best way to resolve this issue would be to collect higher frequency data, which would provide greater information to assess how marketing decisions respond to stochastic production levels. Unfortunately, this is not feasible for this study, for which only annual data is available.

Given the sequential structure of the decision process, the problem must be solved recursively, similar to a dynamic programming model. First the determinants of producer's market participation are solved for, assuming the production decision has already taken place. Then, the determinants of smallholder's production decision can be derived based on their *a priori* knowledge of the factors which will subsequently influence marketing. For example, the realized production and amount of dairy to be bought or sold will depend on the amount of rainfall in a given year.[6] Thus, actual rainfall affects household's participation in the market, but when a farmer is making production decisions they can only form an expectation of what rainfall may be.

4.1 Structural Model

To more formally derive this model, start with the (second-stage) market participation decision, and follow the traditional model described by Key et. al. (and later Bellmare

[6] Rainfall affects the amount of fodder or grain available as an input for dairy producers to feed cattle.

and Barrett, Holloway et. al., and others). They posit a representative agent maximizes their utility (4.1) subject to equations (4.2) through (4.5):

$$max\ u(c) \tag{4.1}$$

$$\sum_{j=1}^{N}\left[\left(p_j^m - t_{pj}^s\left(z_t^s\right)\delta_j^s\right) + \left(p_j^m + t_{pj}^b\left(z_t^b\right)\delta_j^b\right)\right]m_j$$
$$- t_{fj}^s\left(z_t^s\right)\delta_j^s - t_{fj}^b\left(z_t^b\right)\delta_j^b + T = 0 \tag{4.2}$$

$$q_j - n_j + A_j - m_j - c_j = 0, \quad j = 1,\ldots,J \tag{4.3}$$

$$G\left(q, n, ; z_q\right) = 0 \tag{4.5}$$

$$c_j, q_j, n_j \geq 0 \tag{4.6}$$

In equation (4.1) u is the agent's utility as a function of a vector of their consumption, c. Equation (4.2), the budget constraint, is where the role of transaction costs is introduced. Here p_j^m is the market price of good j, and m_j is the amount of that good marketed, which is positive for sellers and negative for buyers. The agent's role in the market is represented by the two indicator functions: δ_j^s is 1 for sellers of good j and 0 otherwise, and δ_j^b is 1 for buyers of good j and 0 otherwise. Note an additional important condition:

$$\delta_j^s + \delta_j^b \leq 1 \tag{4.7}$$

which establishes *m* as the net quantity marketed by stating that a household cannot be both buyer and seller in the same period. In this equation, the proportional transaction costs for sellers of good j, t^s_{pj}, and fixed transaction costs for sellers of good j, t^s_{fj}, effectively change the price they receive and thus their behavior in the market. Similarly, the proportional transaction costs for buyers of good j, t^b_{pj}, and fixed transaction costs for buyers of good j, t^b_{fj}, effectively change the price they pay and thus their behavior in the market. However, as the authors point out, these transaction costs are largely unobserved in survey data, and are thus represented as functions of more readily enumerable factors explaining them, z^s_t and z^b_t respectively. One of the reasons transaction costs are an important element in this model is their role in explaining autarkic behavior of producers, as described in Appendix A. The inclusion of non-market transfers, *T*, which can be positive or negative, completes this constraint.

Equation (4.3) is a feasibility constraint which indicates that for any good j, the amount consumed, c_j, the amount marketed m_j, and the amount used as an input, n_j, cannot exceed the amount produced, q_j, and the endowment A_j. In the case of dairy, this model will impose the simplifying assumption that observations enter each period without an endowment, and that dairy products are not used as inputs. As will be described below, this does not imply that households enter each period without endowments of factors related to dairy production, but that they begin (and end) each period without stocks of dairy products.

Equation (4.4) describes the relationship between inputs, n_j, and outputs through the production technology *G*, and considering other supply shifters, z_q. Recall, at this stage, production decisions regarding n_j have already been made, so we take n_j as given. Traditionally, specifications of z_q have been limited primarily to community level characteristics (such as share of local farmers using fertilizer or hybrid seeds), and endowments over which the household has little control (such as age of household head or the amount of land cultivated) (Key et. al., 2000; and others). The importance of these factors is not disputed. However, one could argue that other important elements may exist in z_q. First, the household's past investments play an important role in the production equation. When investment values have been included in prior analyses, such as livestock assets and trucks in Key et. al. (2000) or total value of assets in Bellemare and Barrett (2007), they have done so contemporaneously. This may arguably lead to biased results as they could be considered endogenous.[7] That is, in a given time period does one own livestock assets because they participate in a given market, or do they participate in the market because they own livestock assets? In reality, the causality likely flows both ways, but this problem can be mitigated by including the lagged investments made by a household.

Secondly, it is important at this stage to separate out variables which would shock the stochastic production process, but which would not be known to producers when

[7] Some household investments, such as trucks in Key et. al., have (correctly) been included as a factor explaining transaction costs rather than a supply shifter. Nevertheless, the argument with respect to being contemporaneously endogenous remains valid.

production decisions are made. The early work of Goetz best addresses these exogenous shocks by acknowledging that quantities produced and consumed are random variables in his structural model, but studies which have followed downplay this fact. Fortunately, with sufficient data we can include realized values of these shocks in the model as observed determinants.

Thus, we can substitute the following constraint, and instead of (4.4), we have:

$$G(q, n,; Z_q) = 0 \tag{4.8}$$

Where $Z_q = \left(z_q^{cc}, z_q^{hi}, z_q^{us}\right)$, and z_q^{cc} are supply shifters associated with community characteristics, as in previous studies, z_q^{hi} are *lagged* household investments, and z_q^{us} are unknown shocks to production, revealed to the farmer after production decisions are made, yet prior to final marketing decisions. Although the household investments in this model are lagged at least one survey period to circumvent endogeneity concerns, for clarity of notation individual household and time period subscripts have been left out of this description of the model.

Indirect utility functions are derived from this optimization, leading to market quantity functions, as in Key et. al., and arrive at the decision rules:

$$m_s = (q - c) = m_s\left(z_t^s, z_q^{cc}, z_q^{hi}, z_q^{us}, n, p^m, T, A\right) \text{,for net sellers,} \tag{4.9}$$

and

$$m_b = (c - q) = m_b\left(z_t^b, z_q^{cc}, z_q^{hi}, z_q^{us}, n, p^m, T, A\right), \text{ for net buyers (4.10)}$$

Notice that, like Bellemare and Barrett and Key et. al., empirically we will separate positive and negative values of the marketed quantity into two non-negative variables for net purchases, m_b, and net sales, m_s. Allowing quantities to be determined by separate processes makes this model more flexible than if we were to use a switching regression, as in Goetz. Similar factors will define threshold quantities which determine whether a household is a net buyer, remains autarkic, or is a net seller. That is, the factors determining whether and in what capacity a producer participates in the market can be thought of as the same as those which determine how much they buy or sell.[8]

There is also a key difference between equations (4.9) and (4.10) and the analogous equations from Key et. al.. That study asserts that factors of production (n) are solvable as a function of prices and production shifters, and thus substitutes inputs out of the quantity equations. To some extent, we agree, as will be illustrated below.[9] This substitution, however, has two drawbacks. First, some of the production shifters outlined in that study (e.g. age of household head, household asset holdings, access to credit, etc.) may have an independent effect on commodity market participation, as discussed when we disaggregated shifters in equation (4.8) above. Including only the shifters, then, in a reduced form would confound their effects on input demand and market participation.

[8] Although not emphasized here, a key theoretical consideration is that fixed transaction costs play a role in the decision on whether to participate in the market, but not on the quantities marketed. However, since these costs are being represented by the factors determining them, which are ultimately treated as the same for both, ignoring this theoretical consideration here has no empirical implications.

[9] Some of the variables described as production shifters in Key et. al. are characterized as either community characteristics (e.g. access to credit) or household characteristics (e.g. age of household head and value of household assets), which appear in the latent model for input demand in equation (4.11).

Secondly, substituting for *n* downplays the fact that it is a vector that includes a wide rage of inputs and alternative technologies. For example, Kenyan dairy farmers can feed by either grazing cattle or using a zero-graze system. They can use (more productive) grade cows or indigenous species. Moreover, several sources contend that successful commercialization hinges on the choice of technology (World Bank, 2008; Kijima et. al. 2006). This hypothesis cannot be tested if *n* is substituted out of the market quantity equations. For these reasons the elements of *n* are not substituted out of market quantity equations, and rather treated as predetermined (and thus contemporaneously exogenous) farmer decisions.

With this established, we turn to the first stage production decision. This *a priori* decision is solved recursively, where farmers are now maximizing expected utility (4.11), subject to the uncertain objective function for market participation (4.12):

$$\max E[u(c)] \tag{4.11}$$

$$D[m] = D[m(z_t, z_q^{cc}, z_q^{hi}, z_q^{us}, n, p^m, T, A)] \tag{4.12}$$

where z_t summarizes factors determining transaction costs for buyers and sellers. In other words, farmers decide on input demand (i.e. production decisions) considering the distribution of possible marketing outcomes (e.g. expected value and, assuming farmers may be risk averse, higher moments). In practice, the determinants of the production decision will be a function of the distribution of marketing determinants themselves. That is, to derive the decision rule for production, we need only consider what the farmer knows about marketing determinants before production takes place.

First of all, the factors affecting transaction costs, z_t^s and z_t^b, could reasonably be considered constant within a given year, and thus considered known to farmers making production decisions. For example, it is unlikely that the distance to a market, or whether there is a private trader in the village would be unknown to the farmer making production decisions. This is also true of community characteristics, z_q^{cc}, such as distance to a veterinarian, whether there is a credit or dairy cooperative, and so on. Lagged household investments, z_q^{hi}, are obviously known to farmers *a priori*.

Resources allocated to production, *n*, are a determinant in the marketing stage, but a decision variable in the production stage. The decision of how to allocate resources for production is theoretically simultaneous. So, obviously, *n* should not be included in the first production decision equation.

The next vector, which by definition is unknown to the farmer prior to production decisions, is the unknown shocks to production, z_q^{us}. Rather than ignore the fact that this will eventually be a determinant in marketing decisions, smallholders may develop some expected value of that shock on which they can base their production decision. Also, since we have implicitly assumed a concave (risk averse) utility function, higher moments of the distribution, such as the variance, will also determine production. This distribution can be written $D\left(z_q^{us}\right)$. The same is true for any prices that will effect marketing decisions (e.g. input, output, and substitute prices), where production is de

ru

n

v

l

w

a

e

decided based on the distribution of market prices, $D(p^m)$. In summary, the decision rules for production decisions can be written:

$$n = n\left(z_t, z_q^{cc}, z_q^{hi}, D\left(z_q^{us}\right), D\left(p^m\right), T, A\right) \tag{4.13}$$

$$w_1 = w_1\left(z_t, z_q^{cc}, z_q^{hi}, D\left(z_q^{us}\right), D\left(p^m\right), T, A\right) \tag{4.14}$$

$$w_1 = 0, \text{ if } n = 0 \quad \text{and} \quad w_1 = 1, \text{ if } n > 0 \tag{4.15}$$

where w_1 is the binary indicator of whether the farmer decides to produce. Although n is also decided at this point, those determinants are not the focus of this study. The empirical model will only include the binary indicator.

5. METHODS

Methods for estimating market participation models have existed in the literature since the early 1990's. These analyze the household unit with respect to agricultural markets, usually employing a "two-tiered" model, otherwise known as a "double hurdle" model. Double hurdles were first introduced as a class of models by Cragg (1971) as a more flexible alternative to tobit models. Goetz (1992), often recognized for his early work in this class of models, analyzes market participation by first separating producing households into participants (buyers and sellers), and non-participants (that is, households which consume only and all of their own production) using probit regression in the first stage. Then, in the second stage, the quantities bought and sold are analyzed along the real line of production less consumption, where negative values denote net buyers and positive values denote net sellers, with a switching regression.[10]

In more recent work Bellemare and Barrett (2006) allow for the consideration of buyers and sellers of livestock separately by first segregating producers into buyers, autarkic, and sellers with ordered probit regression in the first tier of their analysis. Then, in the second tier, truncated normal regressions are estimated for net quantities bought and sold separately for net buyers and net sellers respectively. There are a number of other prominent models between the early work of Goetz and the more recent work of Bellemare and Barrett, although the latter seems to be the most flexible to date.[11] One major drawback to all of these approaches, as described in the previous section, is that

[10] This is called a switching regression because as observations move from negative space to positive space they "switch" from being buyers to sellers.
[11] See Key et. al., Holloway et. al., and others.

each is reliant on the implicit assumption that all observations (and therefore all members of the population analyzed) are producers of the good of interest.

In some cases such two-tiered analysis may indeed be appropriate. For instance, if one is interested in staple grain production of the rural farm population of a developing country, it is fairly reasonable to accept the assumption that all observations are at least producing some. However, in analyzing higher return growth markets (like Kenyan dairy) traditional models are not appropriate if a portion of the population are not producers.

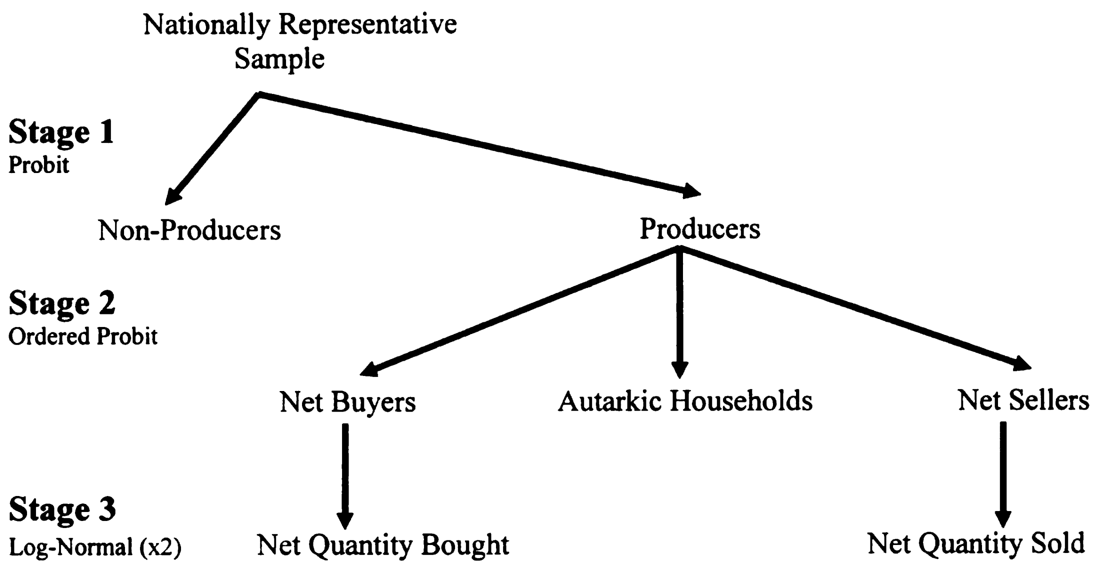
The common solution is to circumvent the issue by focusing on a sub-population of producers. Bellemare and Barrett, for example, study the determinants of livestock market participation for a population of "pastoralists in… Northern Kenya and Southern Ethiopia," within which a random sample is drawn. Holloway et. al. examine Ethiopia's Dairy market participation, but focus only on dairy producers in their sample. While this solves the data problem and allows the research to draw conclusions on determinants within that sub-population, it is not possible to extrapolate policy relevant implications to a national level. Moreover, such a model does not allow the researcher to identify determinants of production itself, the necessary precursor to any market related decisions.

In this analysis of Kenya's dairy market a nationally representative perspective will be maintained by adding an additional stage of analysis, thus employing a 3-stage, or triple hurdle model. Starting with a nationally representative sample, the first stage distinguishes producers from non-producers using probit analysis. In the second stage, similar to the first stage of Bellemare and Barrett, an ordered probit is used to identify factors within producing households which determine whether they are net buyers, autarkic households, or net sellers. Finally, in the third stage, the determinants of buyer and seller quantities are identified in separated log-Normal regressions, which are appropriate given the truncated nature of the dependant variables.

The triple hurdle model also has a major advantage over double hurdle models in general. Double hurdles are useful because they allow a subset of the data to "pile-up" at some value without causing bias in estimating the determinants of the continuous variable for the remaining sample. In many cases this is used for "corner solutions" where optimization behavior results in a zero value (e.g. for charitable contributions, as in Wooldridge, 2002). Double hurdles can also be used in estimating "selection models," where the continuous variable is unobserved for some subset of the sample. Thus far, double hurdles have been able to allow for *either* censored zeros *or* selected zeros. The 3-stage model, on the other hand, allows for the simultaneous existence of both types of zeros. In the case of dairy, that is, the model allows for the market participation variables to be zero either because the household selected themselves out of production altogether, or because the producing household's optimizing market participation is autarkic. Moreover, for any given household, the triple hurdle model will predict the probability that the household is a non-producer and an autarkic producer separately.

The remainder of this section will be used to develop the likelihood function for the 3-stage market participation model just described which is illustrated in Figure 2.

Figure 2. Graphic Illustration of the Three Tiered Market Participation Model



There are two methods available for estimating double hurdle models: Heckit and standard Full Maximum Likelihood. Heckit estimation, first introduced by Heckman (1976) and employed by Bellemare and Barrett (2006), controls (and tests) for selection bias by using a predicted inverse Mill's ratio (IMR) generated using first stage results as a regressor in the second stage. Note that this method is not appropriate in corner solution models under the assumption of joint normality of error terms (Wooldridge, forthcoming). The FML method, on the other hand, maximizes the likelihood function which describes the full probability distribution of all stages.

In this study, both methods could arguably be employed, since the first two stages can be thought of as a selection double hurdle, while the second and third stages represent a combination of corner solution double hurdles.

To derive a likelihood function, we begin in the first stage where households are identified according to whether they are producers or not using probit analysis. To simplify notation from the conceptual framework, let y_1 be the level of milk production, x the vector of all variables thought to explain production and market participation behavior throughout the model, and w_1 is a binary indicator function such that:

$$w_1 = 1[y_1 > 0] \tag{5.1}$$

$$w_1 = 0[y_1 = 0] \tag{5.2}$$

Then, following the standard probit method, we assume:

$$\Pr(w_1 = 1 \mid x_1, \gamma) = \Phi(x_1\gamma) \tag{5.3}$$

$$\Pr(w_1 = 0 \mid x_1, \gamma) = 1 - \Phi(x_1\gamma) \tag{5.4}$$

Where Φ is the standard Normal cumulative distribution function, x_1 are the independent variables thought to determine production, and γ is a vector of parameters to be estimated. Thus, the full distribution of w_1 is:

$$f(w_1 \mid x_1) = [1 - \Phi(x_1\gamma)]^{1[w1=0]}[\Phi(x_1\gamma)]^{1[w1=1]} \tag{5.5}$$

Now, focusing on the second stage, we define y_2 as the level of milk consumption and w_2 the ordered indicator function such that:

$$w_2 = 0[y_1 - y_2 < 0] \tag{5.6}$$

$$w_2 = 1[y_1 - y_2 = 0] \tag{5.7}$$

$$w_2 = 2[y_1 - y_2 > 0] \tag{5.8}$$

In words, w_2 is zero for producing households that are net buyers of milk, w_2 is one for autarkic producing households, and w_2 is two for producing households that are net sellers of milk. Then, following the ordered probit model, define the latent variable w_2^*:

$$w_2^* = x_2\beta + e \qquad e|x_2 \sim \text{Normal}(0,1) \tag{5.9}$$

Let $\alpha_1 < \alpha_2$ be unknown threshold parameters defined such that:

$$w_2 = 0 \quad \text{if } w_2^* < \alpha_1 \tag{5.10}$$

$$w_2 = 1 \quad \text{if } \alpha_1 < w_2^* < \alpha_2 \tag{5.11}$$

$$w_2 = 2 \quad \text{if } w_2^* > \alpha_2 \tag{5.12}$$

Then, letting x_2 be the independent variables explaining market participation:

$$\Pr(w_2 = 0 \mid x_2, \alpha, \beta) = \Pr(w_2^* \le \alpha_1 \mid x_2) = \Phi(\alpha_1 - x_2\beta) \tag{5.13}$$

$$\Pr(w_2 = 1 \mid x_2, \alpha, \beta) = \Phi(\alpha_2 - x_2\beta) - \Phi(\alpha_1 - x_2\beta) \tag{5.14}$$

$$\Pr(w_2 = 2 \mid x_2, \alpha, \beta) = 1 - \Phi(\alpha_2 - x_2\beta) \tag{5.15}$$

Thus, the distribution of w_2 is the ordered probit:

$$f(w_2 \mid x_2) = [\Phi(\alpha_1 - x_2\beta)]^{1[w2=0]}$$
$$* [\Phi(\alpha_2 - x_2\beta) - \Phi(\alpha_1 - x_2\beta)]^{1[w2=1]} [1 - \Phi(\alpha_2 - x_2\beta)]^{1[w2=2]} \quad (5.16)$$

Finally, in the third stage, let y_3 be defined as the net purchases for net buyers, while y_4 is the net sales for the net sellers. Mathematically:

$y_3 = y_2 - y_1$, if $y_2 > y_1$, and is undefined otherwise (5.17)

$y_4 = y_1 - y_2$, if $y_1 > y_2$, and is undefined otherwise (5.18)

As stated above, each of these random variables is assumed to be log-Normal, so, letting x_3 be the independent variables explaining net purchases, and x_4 those explaining net sales, the individual distribution of each can be written:

$$f(y_3 \mid x_3, \delta_3) = \phi[\{\log(y_3) - x_3\delta_3\}/\sigma_3]/(y_3\sigma_3) \quad (5.19)$$

and,

$$f(y_4 \mid x_4, \delta_4) = \phi[\{\log(y_4) - x_4\delta_4\}/\sigma_4]/(y_4\sigma_4) \quad (5.20)$$

Where ϕ is the standard normal probability density function, and σ_j is the standard deviation of the random variable y_j. It should be noted here that, as with the double hurdle models, there are no restriction regarding the elements of x_1, x_2, x_3, and x_4 (i.e. they can be the same or different explanatory variables in each stage). That said, as discussed during the derivation of the structural model for this study, the elements of

x_2, x_3, and x_4 will be the same. Nevertheless, we can finalize the derivation of our likelihood function in the more general case, where each vector is treated separately. Finally, defining θ as the vector of all the above described parameters, and using exponential indicator functions, the joint distribution function for w_1, w_2, y_3, y_4 can be written:

$$f(w_1, w_2, y_3, y_4 \mid x, \theta) = [1 - \Phi(x_1\gamma)]^{1[w1=0]} \tag{5.21}$$

$$* \left[\Phi(x_1\gamma) \left\{ \begin{array}{l} \left[\Phi(\alpha_1 - x_2\beta) \dfrac{\phi[\{\log(y_3) - x_4\delta_3\}/\sigma_3]}{y_3\sigma_3} \right]^{1[w2=0]} \\ \left[\Phi(\alpha_2 - x_2\beta) - \Phi(\alpha_1 - x_2\beta) \right]^{1[w2=1]} \\ \left[(1 - \Phi(\alpha_2 - x_2\beta)) \dfrac{\phi[\{\log(y_4) - x_4\delta_4\}/\sigma_4]}{y_4\sigma_4} \right]^{1[w2=2]} \end{array} \right\} \right]^{1[w1=1]}$$

Thus, for any given observation i, the log-likelihood function is:

$$
\begin{aligned}
\ell_i(\theta) = {} & 1[w_{1i}=0]\log[1-\Phi(x_{1i}\gamma)] + 1[w_{1i}=1] \\
& + 1[w_{1i}=1]\left[\begin{array}{l}
\log[\Phi(x_{1i}\gamma)] + 1[w_{2i}=0]\left(\begin{array}{l}\log[\Phi(\alpha_1 - x_{2i}\beta)] - \log(y_{3i}) - \frac{1}{2}\log(\sigma_3^2) \\ -\frac{1}{2}\log(2\pi) - \frac{1}{2}\left[\frac{\{\log(y_{3i}) - x_{3i}\delta_3\}^2}{\sigma_3^2}\right]\end{array}\right) \\
+ 1[w_{2i}=1](\log[\Phi(\alpha_2 - x_{2i}\beta) - \Phi(\alpha_1 - x_{2i}\beta)]) \\
+ 1[w_{2i}=2]\left(\begin{array}{l}\log[1-\Phi(\alpha_2 - x_{2i}\beta)] - \log(y_{4i}) - \frac{1}{2}\log(\sigma_4^2) \\ -\frac{1}{2}\log(2\pi) - \frac{1}{2}\left[\frac{\{\log(y_{4i}) - x_{4i}\delta_4\}^2}{\sigma_4^2}\right]\end{array}\right)
\end{array}\right]
\end{aligned}
\tag{5.22}
$$

Although this study will maximize equation (5.22) simultaneously, it should be noted that, due to the separability of the likelihood function, the MLE of γ can be obtained by a probit regression of w_1 on x_1, the MLE of β can be obtained from an ordered probit regression of w_2 on x_2 using only observations for whom w_1=1, the MLE of δ_3 can be obtained by regressing y_3 on x_3 using only those observations where w_2=0, and the MLE of δ_4 can be obtained by regressing y_4 on x_4 using only those observations where w_2=2. That said, separability in estimation should not be mistaken for separability in interpretation. That is, while various parameter vectors could be obtained separately, the overall effect of a given independent variable can only be determined when considering all parameter vectors together. This will be demonstrated below.

Results from maximizing the likelihood function in equation (5.21) can be used to predict the following "unconditional" probabilities for any given observation:[12]

$$\Pr_i(\text{Not producing}) = \Pr(w_{1i} = 0 \mid x) = 1 - \Phi(x_{1i}\gamma) \tag{5.23}$$

$$\Pr_i(\text{Net buyer}) = \Pr(w_{1i} = 1, w_{2i} = 0 \mid x) = \Phi(x_{1i}\gamma)\Phi(\alpha_1 - x_{2i}\beta) \tag{5.24}$$

$$\begin{aligned}\Pr_i(\text{Autarkic}) = {} & \Pr(w_{1i} = 1, w_{2i} = 1 \mid x) \\ & = \Phi(x_{1i}\gamma)\left(\Phi(\alpha_2 - x_{2i}\beta) - \Phi(\alpha_1 - x_{2i}\beta)\right)\end{aligned} \tag{5.25}$$

$$\begin{aligned}\Pr_i(\text{Net seller}) = {} & \Pr(w_{1i} = 1, w_{2i} = 2 \mid x) \\ & = \Phi(x_{1i}\gamma)\left(1 - \Phi(\alpha_2 - x_{2i}\beta)\right)\end{aligned} \tag{5.26}$$

---

[12] "Unconditional" is a bit of a misnomer, since all predicted probabilities are conditional on the independent variables. Also, since this study uses a balanced panel, estimates are conditional on households being intransient. Here, unconditional is only meant to imply that the probabilities (and later expected values) are for any given observation, not conditional on any of the dependent variables (production or market participation) taking a specific value.

Once again, one of the major benefits of the model is the ability to treat non-producers and autarkic producers separately, as shown in equations (5.23) and (5.25). Results can also be used to predict the unconditional expected values of net sales and net purchases for any given observation:

E_i(Net Purchases) =

$$E(y_{3i} \mid x) = \Phi(x_{1i}\gamma)\Phi(\alpha_1 - x_{2i}\beta)\exp\left(x_{3i}\delta_3 + \sigma_3^2/2\right) \tag{5.27}$$

E_i(Net Sales) =

$$E(y_{4i} \mid x) = \Phi(x_{1i}\gamma)(1 - \Phi(\alpha_2 - x_{2i}\beta))\exp\left(x_{4i}\delta_4 + \sigma_4^2/2\right) \tag{5.28}$$

The ability to predict expected values in equations 5.27 and 5.28, without conditioning on the observation being a producer is only possible with a triple hurdle model, and allows us to extrapolate implications to a national level. Finally, using the product rule, the partial effect of any continuous explanatory variable (x_k) on the unconditional expected values can be derived. Since the market participation of net sellers is of primary interest to this study, we'll focus on the partial effect on the unconditional value of net sales:

$$\begin{aligned} \frac{\partial E(y_{4i} \mid x)}{\partial x_k} &= \gamma_k \phi(x_{1i}\gamma)(\Phi(x_{2i}\beta - \alpha_2))\exp\left(x_{4i}\delta_4 + \sigma_4^2/2\right) \\ &+ \beta_k \Phi(x_{1i}\gamma)(\phi(x_{2i}\beta - \alpha_2))\exp\left(x_{4i}\delta_4 + \sigma_4^2/2\right) \\ &+ \delta_{4k} \Phi(x_{1i}\gamma)(\Phi(x_{2i}\beta - \alpha_2))\exp\left(x_{4i}\delta_4 + \sigma_4^2/2\right) \end{aligned} \tag{5.29}$$

Notice, equation (5.29) is derived under the assumption that the explanatory variable of interest is an element of each vector ($x_k \in x_1, x_2, x_4$), or that the explanatory variable is included in each stage of the 3-stage model. If that is not the case, the partial effect can be derived from (5.29) simply by setting the parameter from the stage where it is omitted to zero. For example, if the explanatory variable is not in the production decision stage ($x_k \notin x_1$), then we know $\gamma_k = 0$, which means the first of the three RHS terms in (5.29) goes to zero. In that case, the partial effect becomes:

$$\frac{\partial E(y_{4i} \mid x)}{\partial x_k} = \beta_k \Phi(x_{1i}\gamma)(\phi(x_{2i}\beta - \alpha_2))\exp\left(x_{4i}\delta_4 + \sigma_4^2/2\right)$$
$$+ \delta_{4k}\Phi(x_{1i}\gamma)(\Phi(x_{2i}\beta - \alpha_2))\exp\left(x_{4i}\delta_4 + \sigma_4^2/2\right) \tag{5.30}$$

From equation 5.30 we can see that even if the explanatory variable is not in the first tier, its partial effect on the unconditional expected value of net sales depends on parameters and explanatory variables from every tier. That is, even though ($x_k \notin x_1$), the partial effect is still a function of the vectors x_1 and γ. This returns us to the point that, although there is separability in estimation, the overall effect of a given variable can only be understood when we consider all parameter estimates simultaneously.

6. VARIABLE CONSTRUCTION

6.1 Community Characteristics and Transaction Cost Determinants

Population density will be included in units of people per square mile. The data, collected at the village level in 1997 only, must be treated as a time constant. Access (or lack thereof) to education will be measured by the prevalence of household heads without any formal education at the division level (actual level of education attained will be controlled for separately and discussed with other household characteristics). An admitted caveat to this approach is the implicit assumption that the availability of education did not change much over time; the household heads are of varying ages, thus would have gone to school at different times. Nevertheless, a prevalence ratio is arguable the best available measure of the accessibility of education. There are also dummy variables for whether the household's area receives a short rainy season and for the 2007 survey year. Zone-level time-varying and fixed effects will also be controlled for.

Expected rainfall is computed as the 7-year average main season rainfall in millimeters. The 7-year lag is the maximum feasible with available data. The variance of rainfall over the same period will also be included to control for the varying reliability of rainfall expectations. Positive and negative shocks to expected rainfall will be controlled for separately in the market participation stages of the model. As discussed by Hoddinott (2006), separating positive and negative shocks allows more flexibility in the model. Both shocks are measured as the absolute value of the difference between actual and expected rainfall. Obviously, in a given year only one type of shock can affect the household (i.e. either the positive or negative shock variable will be zero).

Market access as a transaction cost will be controlled for using the distance to the nearest motorable road, measured in kilometers. Access to veterinary services and electricity will be controlled for by using the distance to the nearest facility in kilometers. While these distances are community characteristics in the sense that they are determined exogenously, data were collected at the household level.

Another key factor is the farmer's access to a purchasing enterprise, other than those for informal household-to-household transactions. Dummy variables will be included for whether the KCC, another dairy cooperative, a formally licensed private trader, or an informal private trader is active in the area (four dummy variables). That is, if any household in the village is selling to one of these institutions, the dummy variable will be 1 and 0 otherwise. These variables will be of particular interest, given the current policy debate over the limited number of purchasing licenses in Kenya, and the importance of private traders (World Bank 2008). A similar dummy variable would be ideal for controlling for access to credit. However, at least some form of credit is available in almost all villages in a given year, so this approach would not provide enough variation between areas to generate a precise estimate of its effect. Rather, the prevalence of credit in the area will be measured as the share of households receiving credit at the division level.

6.2 Household Characteristics and Investments

Household characteristics controlled for include the age of the household head and a dummy variable for female headed households. The model will also control for the household's number of adult equivalents, which will be calculated using the World Bank's age and gender based scale.

Lagged household investments will include the number of acres of land they owned as well as the total real value (in 10,000 2007 Ksh units) of the households other productive assets, all in the previous period.[13] In addition to the availability of education, discussed above, the actual investment in education will be controlled for by including a set of 4 dummy variables for the education of the household head: 1 to 4 years, 5 to 8 years, 9 to 12 years, and more than 12 years or some college. The effects of these can be interpreted as a comparison to no formal education, the effects of which will be subsumed into the intercept term.

The model will include a dummy variable for whether the household experienced the death of a prime-aged adult (15-59 years) between interviews (i.e. over approximately the past three years). A variable will also be included to control for the number of chronically ill people in the household during the production year. It will be assumed that this is an unanticipated shock, and so will be omitted from the production decision stage.

The household level of production factors, as described by the structural model, are implicitly endogenous (and thus not included as explanatory) in the first stage, but taken as "state" or "pre-determined" variables in the second and third stages, as in a dynamic programming model. The assumption one needs to make in order for this to be logical is to say that when a farmer is buying or selling milk, they cannot change the number of cows used to produce it. Under the sequential decision making process assumed by the structural model, this is reasonable. The production factors included are the number of

[13] Productive assets include: ploughs (tractor and animal traction), cart, trailer, tractor, cars, trucks, spray pump, irrigation equipment, water tanks, stores, wheelbarrow, combine harvester, donkey, bulls, chickens, goats, sheep, calves, cows, pigs, turkeys, and ducks.

grade cows owned during the year, the number of acres farmed (rented or owned), as well as a dummy variable for whether the household used zero-grazing technology, which is thought to promote higher yields. Though conceptually pre-determined, the number of grade cows and zero-grazing dummy variables will be lagged one period to mitigate endogeneity concerns.

6.3 Prices

The real market prices included are for milk (2007 Ksh/liter), tomatoes (2007 Ksh/kg), maize grain (2007 Ksh/kg), and DAP fertilizer (2007 Ksh/50 kg bag). Tomatoes, one of the most popular horticultural crops in the sample, are included to represent the price of potential product from an alternative use of resources. Maize grain is another alternative product for households, but also an input as feed in dairy production. For that reason, DAP can also be thought of as an input for dairy production, and its price an input price. For crops and dairy we would ideally use different prices for producers and consumers, but such data are not consistently available. Rather, we must use producer prices, allowing the difference between them and the consumer prices to be captured by the included transaction cost variables.

Of course, when production decisions are being made the actual market price of outputs will not be known, so the first stage model should include price expectations (and higher moments). This is particularly true of dairy, which requires a high initial investment compared to crops, and so decisions are likely made based on price expectations several periods into the future. With the given data, the best indicator available is actual current prices. As a proxy for price expectations, of course, this implies perfect foresight for one period. However, the only alternative would be using a three year lagged price, which would imply a very naïve expectation. Unfortunately, we lack the time series data that would be necessary to test these assumptions against each other for all of the relevant prices. Thus, barring the collection of substantially more data requiring recall responses dating back more than a decade, assuming perfect foresight for one period is the best feasible option.

It should be noted that assuming current prices represent expected future prices has few practical implications for key policy variables such as distance to veterinarian services and access to purchasing enterprises. This is explored further in Appendix B, but in general the coefficients of key variables change very little in magnitude or statistical significance in the first stage results when price expectations are dropped altogether, and do not change at all in the second and third stage results due to the separability of the likelihood function in equation (5.22).

A summary and some descriptive statistics for the variables used in the model can be found in Table 1, which show various percentiles and means of continuous variables, and Table 2, which shows shares of households with positive values for dummy variables by year.

Table 1: Distribution of Continuous Explanatory Variables

Explanatory Variable	Percentile 10	25	50	75	90	Mean
Regional Characteristics						
Population Density by Division (People/squared mile)	118.00	169.00	314.00	595.00	1242.00	481.98
Distance to motorable road in km	.03	.10	.40	1.00	2.00	.79
Distance to vet services(km)	.50	1.00	3.00	5.00	8.00	4.15
Distance to electricity (km)	.10	.88	2.00	4.00	9.00	4.17
Division % of household heads w/o education	.08	.14	.17	.22	.33	.20
Share hh in village receiving credit	.07	.17	.33	.64	.93	.42
Lagged number of grade cows	.00	.00	.00	.00	2.00	.44
Lagged acres owned	.65	1.25	2.40	4.34	6.95	3.50
Lagged asset value (,000 2007 Ksh)	5.65	23.25	55.88	128.84	288.92	142.48
Past 7 year average Main Season Rainfall (mm)	281.22	377.49	579.15	688.49	764.74	538.98
Past 7 year variance Main Season Rainfall (mm^2)	7449.8	15030.2	23025.8	38707.5	58491.1	38718.0
Positive rain shock (mm) vs. 7 year mean	.00	.00	107.01	284.23	374.48	152.43
Negative rain shock (mm) vs. 7 year mean	.00	.00	.00	29.08	112.25	28.53
Household Characteristics						
Number of chronically ill members during year	.00	.00	.00	.00	1.00	.23
Age of Household Head	40.00	48.00	57.00	67.00	75.00	57.58
Adult Equivalents	2.50	3.78	5.52	7.36	9.22	5.74
Main season acres farmed	.76	1.28	2.37	4.14	7.00	3.47
Prices						
Real district price of maize (2007 Ksh/kg)	11.11	12.22	15.02	18.05	18.81	15.12
Real district price of tomatoes (2007 Ksh/kg)	7.66	7.86	9.43	11.01	13.16	9.74
Real district price of milk (2007 Ksh/litre)	15.00	17.00	20.31	27.08	30.00	21.88
Real district price of 50 kg DAP bag (2007 Ksh)	1700.00	1760.32	1800.00	1828.03	1895.73	1802.09

Source: Tegemeo household surveys 2000, 2004, 2007. Rainfall data from Kenyan national weather service climate prediction center as part of the Famine Early Warning System project.

Table 2: Distribution of Binary Explanatory Variables over Time

	Year	
	2004	2007
Explanatory variable	Share of Households	
Household has short rain season	.80	.80
Milk Cooperative in village	.20	.23
Kenya Creamery Company in village	.16	.14
Private processors/traders in village	.31	.29
Informal trader/hawker in village	.42	.46
Did HH experience a Prime Age Death (15 to 59)	.05	.05
female headed household	.20	.24
No education	.20	.19
1 to 4 years	.18	.17
5 to 8 years	.34	.37
9 to 12 years	.20	.20
More than 12 years (or some college)	.07	.07
Use zero grazing technology (lagged)	.12	.14

Source: Tegemeo household surveys 2004, 2007.

7. RESULTS

Results from estimation of the triple hurdle model for dairy market participation in Kenya can be found in Table 3. Column (i) shows results from the first stage, which estimates the determinants of the probability of producing milk. Column (ii) shows results from the second stage, which estimates the determinants of being either a net buying, autarkic, or net selling household, conditional on being a producer. These results are interpreted as in the ordered probit, where positive coefficients imply that as the explanatory variable increases the observation is more likely to be autarkic than a net buyer and more likely to be a net seller than autarkic. Columns (iii) and (iv) show the estimation results for the determinants of net purchase and net sales volumes, conditional on being a producer and net buyer or producer and net seller respectively. Again, due to the non-linear nature of the likelihood function at all tiers of this model, these coefficients are not partial effects. The marginal effect of any given variable for a given observation will depend upon the level of all other explanatory variables for that observation, as in equations (5.29) and (5.30). Nevertheless, the direction and statistical significance of effects can be analyzed using these results. For a closer look at how these factors may affect actual outcomes, we will later conduct a series of simulation analyses using these results and the properties described in equations (5.23) through (5.28).

The first thing to note is the absence of a predicted IMR among the explanatory variables in the second stage. Recall, the first two stages can be considered a selection model, which could be estimated using the Heckman procedure to eliminate selection bias. As

Table 3: Three-Stage Model for Dairy Market Participation in Kenya (MLE)

	Stage 1: Production	Stage 2: Market Participation	Stage 3: Net Purchases	Stage 3: Net Sales
	Probit	Ordered Probit	Log Normal	
	(i)	(ii)	(iii)	(iv)
year = 2007	0.440	-0.066	-0.712	-1.065***
	(1.60)	(0.08)	(0.85)	(2.65)
Control for zone fixed effects[a]	*Yes*	*Yes*	*Yes*	*Yes*
Control for time variant zone effects	*Yes*	*Yes*	*Yes*	*Yes*
Regional Characteristics				
Population density by division (people/mi^2)	-0.000	-0.000	0.000	0.000
	(1.41)	(0.39)	(1.42)	(1.47)
Distance from motorable road (km)	0.017	0.056	0.033	-0.047
	(0.58)	(1.50)	(0.41)	(1.40)
Distance to vet services(km)	0.014**	0.004	-0.005	0.000
	(1.97)	(0.58)	(0.21)	(0.01)
Distance to electricity (km)	-0.009	-0.020***	-0.016	-0.012*
	(1.45)	(3.37)	(1.10)	(1.75)
Household has short rain season	0.006	-0.173	0.548	0.201
	(0.04)	(0.92)	(1.16)	(0.97)
Division share of household heads with no education	-0.605	0.101	-1.287	0.145
	(1.42)	(0.18)	(0.95)	(0.26)
milk cooperative in village	0.094	0.493***	-0.688**	0.268***
	(0.97)	(4.00)	(2.23)	(2.62)
Kenya creamery company in village	0.188*	0.336***	0.526	0.111
	(1.78)	(2.63)	(1.64)	(1.05)
private processors/traders in village	0.448***	0.301***	-0.011	0.323***
	(5.15)	(2.94)	(0.05)	(3.50)
informal trader/hawker in village	0.162**	0.180*	-0.237	0.255***
	(2.04)	(1.82)	(1.14)	(2.78)
share hh in village receiving credit	0.306*	0.541***	-0.612	0.328**
	(1.95)	(2.94)	(1.20)	(2.05)
past 7 year average main season rainfall (mm)	-0.000	-0.002	0.001	0.001
	(0.49)	(1.61)	(0.55)	(0.67)
past 7 year variance main season rainfall (mm^2/1000)	0.004***	0.003*	-0.011	0.002*
	(3.42)	(1.86)	(1.56)	(1.67)
Positive rain shock (mm) vs. 7 year mean	-	0.000	-0.002	0.001
	-	(0.11)	(1.35)	(1.63)
negative rain shock (mm) vs. 7 year mean	-	0.001	-0.002	0.002
	-	(0.38)	(0.41)	(1.17)
Household Characteristics				
lagged acres owned	0.059***	0.006	0.002	0.019
	(3.74)	(0.47)	(0.07)	(1.08)
lagged asset value (10,000 2007 Ksh)	0.016**	0.004*	0.006*	0.000
	(2.16)	(1.77)	(1.86)	(0.19)
Death of prime aged adult over the past 3 years[b]	-0.124	0.050	0.012	-0.053
	(0.95)	(0.33)	(0.03)	(0.31)
Number of chronically ill household members during year	-	0.049	-0.253	-0.120
	-	(0.73)	(1.33)	(1.63)
female headed household	-0.088	0.183**	0.136	0.021
	(1.04)	(2.03)	(0.56)	(0.22)

Table 3 (continued)

1 to 4 years[a]	-0.002	0.176	0.009	-0.114
	(0.02)	(1.58)	(0.03)	(1.02)
5 to 8 years	0.149	0.142	0.080	-0.053
	(1.40)	(1.32)	(0.28)	(0.46)
9 to 12 years	0.173	0.198	0.302	0.220
	(1.36)	(1.41)	(1.00)	(1.64)
more than 12 years (or some college)	0.314*	0.633***	0.511	0.289**
	(1.88)	(3.60)	(1.17)	(1.97)
age of household head	0.009***	-0.005	0.008	0.004
	(2.88)	(1.48)	(1.02)	(1.38)
adult equivalents	0.054***	-0.018	0.059*	0.011
	(3.74)	(1.25)	(1.82)	(0.83)
lagged number of grade cows	-	0.118***	0.034	0.096***
	-	(3.00)	(0.32)	(3.45)
use zero grazing technology (lagged)	-	0.458***	0.097	0.204*
	-	(3.45)	(0.32)	(1.93)
main season acres farmed	-	0.037**	0.034	0.036**
	-	(2.51)	(1.44)	(2.30)
Prices				
real district price of milk (2007 ksh/litre)	-0.018	0.036*	-0.013	-0.009
	(1.14)	(1.95)	(0.33)	(0.46)
real district price of maize (2007 ksh/kg)	0.117***	0.015	-0.054	-0.016
	(4.54)	(0.50)	(0.69)	(0.41)
real district price of tomatoes (2007 ksh/kg)	0.005	-0.056	-0.170	-0.004
	(0.13)	(1.20)	(1.50)	(0.08)
price of 50 kg dap bag (2007 ksh)	-0.000	-0.001	-0.003	-0.001
	(0.50)	(0.78)	(1.14)	(0.71)
Constant	-3.381*	-	14.411**	7.919***
	(1.78)	-	(2.24)	(3.22)
Ancillary Parameters				
α_1	-	-2.147	-	-
		(0.86)		
α_2	-	-1.673	-	-
		(0.67)		
σ	-	-	1.266***	1.105***
			(19.88)	(34.50)
Observations	2550	1753	291	1257

A comprehensive evaluation of the model's overall performance can be found in Appendix E.
Full results with standard errors, p-values, and 95% confidence intervals can be found in Appendix C.
Notes: * significant at 10%; ** significant at 5%; *** significant at 1%. z statistics robust to heteroskedasticity and autocorrelation in parentheses. (a) Dummy variable for Coastal lowlands and no education are omitted due to collinearity, their effects being subsumed into the intercept terms. (b) Prime aged adults are 15 to 59 years of age.

noted in Wooldridge (2002), when this is done a simple t-test on the coefficient of the predicted IMR can be used to test the null hypothesis of no selection bias. When the IMR is included, the p-value for that test is 0.21, giving evidence that we should not reject the null hypothesis (i.e. the evidence suggests there is no selection bias). Since the IMR is a statistically irrelevant variable, it is omitted from final analysis.

The results on the effects of average rainfall suggest a negligible effect on probability of producing, but that in areas where rainfall expectations are higher on average, households are more likely to be net buyers (significant at 10.7% level). Farmers in such areas are likely producing more crops, getting higher yields, and supplementing their own production with milk purchases. Interestingly, the results also show that the coefficient on rainfall variance is positive and significant in the probability of producing, probability of a producer being a net seller, and in the quantity of net sales. In other words, where rainfall is less consistent and expectations less reliable, farmers opt for dairy production as a steadier source of income than, say, crop income. This may seem counterintuitive, since cattle are either grain- or fodder-fed, but evidently dairy income is relatively insulated from rainfall shocks. Indeed, the rainfall shock variables are generally not significant in either direction on the probability of being a participant or marketing volumes. The only exception is that positive shocks in rainfall have a positive and significant (at 10.3% level) effect on net sales volume among net sellers.

When considering the actual amount of education obtained by the household head (represented by the four dummy variables), all the significant coefficients support the notion that education is positively correlated with production and sale of milk. This is especially true for the "more than 12 years or some college" dummy variable, which is positive and significant in the production, participation, and net sales volume stages. This is probably showing the effects of rural technical college educations, which are included in "some college".

There is also evidence that extension services are an important factor in dairy market participation. Although distance to veterinary services only appears significant in the production decision stage, the number of grade cows previously owned was highly significant in all decision stages for net sellers. This illustrates the importance of artificial insemination services in a farmer's ability to be a net seller of dairy. Furthermore, after controlling for herd size, the use of zero-grazing technology increases the probability of being a net seller, as well as the quantity of net sales.

Distance to electricity does not have a significant effect on the probability of being a producer. On the other hand, as distance to electricity increases for a given producer, they are more likely to be autarkic or net buyer than a net seller. Also, as distance to electricity increases for a given net seller, their quantity of net sales will go down, and these relationships are significant at a 1% and 10% level respectively. The importance of electricity in dairy marketing is probably two-fold. First, electricity is a necessary input in the AI process, and, as evidence already discussed suggests, AI and the use of improved cow species is an important factor in dairy marketing. Second, electricity allows for cold-storage, which prolongs milk's shelf-life, promotes safety, and generally facilitates the sale of dairy products.

Table 4: Average[a] Household Production and Participation Probability Simulations by Credit Prevalence and Distance to Electricity

	"Unconditional"[b] Probability of Being:				
	Non-producer	Producer			Total
		Net Buyer	Autarkic	Net Seller	
State of Nature	(i)	(ii)	(iii)	(iv)	(v)
Share of households with credit and distance to electricity					
95% and 1 km	0.204	0.063	0.076	0.658	1.00
75% and 2 km	0.224	0.077	0.085	0.614	1.00
50% and 4.5 km	0.255	0.101	0.097	0.547	1.00
25% and 10 km	0.296	0.139	0.110	0.455	1.00
Data means	0.262	0.106	0.099	0.533	1.00

Source: Tegemeo household surveys 2000, 2004, 2007, and author's calculations.
Notes: (a) All explanatory variables not described in the "State of Nature" are held constant at the national mean values. (b) "Unconditional" implies results are not conditioned on being a producer (all probabilities are conditional on explanatory variables). For example, "Net Seller" column shows the probability of any given household being a net seller, versus the probability that a given producer is a net seller.

Credit prevalence is a similarly important determinant in all stages of the net sellers production and marketing decisions. As credit becomes more available (prevalence ratio increases) farmers are more likely to produce, more likely to be a net seller, and quantity sold increases. All of these relationships are significant at the 10% level or better.

Once again, the coefficient estimates in Table 3 should not be thought of as partial effects, so the importance of distance to electricity and credit prevalence is more readily quantified using the simulation analysis in Table 4. This type of post-estimation analysis is similar to that described by Long and Freese (2006) for understanding categorical dependant variable models, although here the method has been updated (as described in equations 5.23 to 5.26) to account for the multi-stage model. Table 4 shows the probabilities of a farmer being a non-producer or a net buying, net selling, or autarkic producer for given states of nature vis-a-vis credit prevalence and distance to electricity.

The simulations in Table 4 show that a household where credit is more readily available (95% prevalence) and electricity is close by (1 km), is about 20% likely to be a non-producer of dairy, and 66% likely to be a net selling producer. Conversely, a household in an area where credit is less available (25% prevalence) and far from electricity (10 km) is 46% likely to be a net selling producer and nearly 30% likely not to produce dairy at all. In other words, the relatively enabled households in the first simulation are 45% more likely to be a net seller than those in the latter simulation, and 23% more likely than a household at the current data means.

Table 5 presents further simulation analysis on the effects of credit prevalence and distance to electricity, focusing on the decisions of net sellers. For various described states of nature, column (i) shows the predicted probability of producing, column (ii) shows the predicted probability of being a net seller, given that the household is a producer, column (iii) shows the expected volume of net sales, given that the household is a producer and net seller, and column (iv) shows the expected volume of net sales without condition on being a producer or net seller. The so-called "unconditional" expected values in column (iv) could also be interpreted as the expected per-household net sales for various states of nature.[14] As noted in section 5, the term "unconditional" expected value is meant only to imply that the values are not conditional on the households being a producer or net seller. All predicted values are conditional on the control variables, and, in this case, households being intransient (since this study is using a balanced panel).

[14] A more complete table of simulation results, covering a wider range of states of nature, can be found in Appendix D, Table D1.

Table 5: Average[a] Household Production and Net Sales Probability and Expectation by Credit Prevalence and Distance to Electricity

State of Nature	Probability of Producing (i)	Probability of net seller, given production (ii)	Expected net sales, given net seller (liters) (iii)	"Unconditional"[b] expected net sales (liters) (iv)
Share of households with credit and distance to electricity				
95% and 1 km	0.796	0.826	1566	1030
75% and 2 km	0.776	0.791	1449	889
50% and 4.5 km	0.745	0.734	1295	708
25% and 10 km	0.704	0.647	1115	508
Data means	0.738	0.722	1266	675

Source: Tegemeo household surveys 2000, 2004, 2007, and author's calculations.
Notes: (a) All explanatory variables not described in the "State of Nature" are held constant at the national mean values. (b) "Unconditional" implies results are not conditioned on being a producer or net seller (all probabilities are conditional on explanatory variables). That is, this column shows the expected net sales of any given household, versus the expected net sales of a given net seller.

Table 5 shows that a household where credit is more readily available (95% prevalence) and electricity is close by (1 km) is 80% likely to be a producer, and that these producers are 83% likely to be net sellers. The expected net sales of these net sellers are nearly 1600 liters per year. On the other hand, a household in an area where credit is less available (25% prevalence) and far from electricity (10 km) is only 70% likely to be a producer, and these producers are only about 65% likely to be net sellers. Moreover, these net sellers are expected to have lower net sales volume, at around 1100 liters per year. The total effect of the difference in these two simulations, shown in column (iv) is dramatic. In the first scenario a given farmers unconditional expected sales are 1030 liters per year, compared to less than half of that (508 liters per year) in the latter scenario. In other words, the per household net sales in the enabling environment described in the first simulation are 53% higher than that at the current data means, and 103% higher than that of households in the latter, less enabled simulation.

Table 6: Average Partial Effect (APE) of Distance to Electricity (km) on Unconditional[a] Expected Volume of Net Dairy Sales (liters/hh)

Zones	APE[b] of Distance to Electricity (liters/hh)	Average Distance to electricity (km)
Coastal Lowlands	-8.014	4.1
Eastern Lowlands	-8.480	6.4
Western Lowlands	-5.816	2.5
Western Transitional	-11.079	3.6
High Potential Maize	-36.596	5.7
Western Highlands	-10.598	2.5
Central Highlands	-28.333	0.8
Marginal Rain Shadow	-33.906	18.2
Total	-20.785	4.2

Source: Tegemeo household surveys 2000, 2004, 2007, and authors calculations.
Notes: (a) "Unconditional" implies results are not conditioned on being a producer or net seller (all probabilities are conditional on explanatory variables). (b) Sample APE significant at 1% level using delta-method for variance approximation (z-score = 3.08, p-value=.004)

Table 6 shows the average partial effect (APE) of distance to electricity on the unconditional expected value of net sales, as shown in equation (5.29). The results are disaggregated by agro-ecological zones, and juxtaposed with the zone-averaged current distance to electricity. Notice first that the APE for the entire sample is -20.785, and the mean distance to electricity is 4.2 kilometers (Sample APE is significant at 1% level using delta-method for variance approximation). This implies that if, through increased electrification, the average distance from household to electricity could be brought to within 1 kilometer, the nationwide per household net sales of dairy would increase by nearly 67 liters per year (a 6% increase in current average expected value). Also of note is the fact that the second largest APE is in the Marginal Rain Shadow zone, which has the least access to electricity. Commercial dairy enterprises are in that area, but there is clearly a substantial amount of untapped potential that would be more easily exploited with more access to electricity and the services it enables.

Regression results also indicate that purchasing enterprises are very important to production and market participation. Admittedly, there is some long-term endogeneity surrounding these variables, since production and marketing segments of the dairy supply chain will grow together over time (fortunately, estimation requires only contemporaneous exogeneity to be unbiased and consistent). Nevertheless, these results can illustrate the overall effectiveness of the relationships between dairy farmers and the various marketing channels in promoting smallholder commercialization.

Households in a village where the KCC is active are more likely to produce and be net sellers. These effects are significant at the 10% and 1% levels respectively. The presence of another (non-KCC) milk cooperative does not show a significant effect on the probability of producing, but producers in the area are substantially more likely to be sellers, and sellers are expected to sell more (both effects significant at 1% level). Furthermore, net buyers in such an area are buying significantly less. Private agents, both formal and informal, also have a significant effect on every stage of the production and marketing decisions for net sellers. When either type of private agent is active in the village, farmers are more likely to produce, producers are more likely to be net sellers, and net sellers are expected to sell more.

Once again, these results can be put into more quantifiable context through simulation. Table 7 shows the probabilities of a farmer being a non-producer or a net buying, net selling, or autarkic producer, depending on which type of purchasing enterprise is active in the village. Notice that when no agent is active in the village, the average household is 35% likely not to produce dairy and 39% likely to be a net selling producer. On the other hand, if the KCC is active in the village, the probability of being a non-producer falls to

Table 7: Average[a] Household Production and Participation Probability Simulations

State of Nature	Unconditional[b] Probability of Being: Non-producer (i)	Producer: Net Buyer (ii)	Producer: Autarkic (iii)	Producer: Net Seller (iv)	Total (v)
Marketing Agent in Village					
Milk Co-op	0.317	0.075	0.079	0.529	1.00
KCC	0.284	0.102	0.095	0.519	1.00
Private Trader	0.203	0.119	0.109	0.568	1.00
Informal Trader	0.293	0.127	0.106	0.474	1.00
None	0.351	0.150	0.108	0.391	1.00

Source: Tegemeo household surveys 2000, 2004, 2007, and author's calculations.
Notes: (a) All explanatory variables not described in the "State of Nature" are held constant at the national mean values. (b) "Unconditional" implies results are not conditioned on being a producer (all probabilities are conditional on explanatory variables). For example, "Net Seller" column shows the probability of any given household being a net seller, versus the probability that a given producer is a net seller.

28%, and the probability of being a net selling producer rises to 52%. The largest effect on the probability of being a producer and net seller is from private traders. When these agents are in the area, the average household is only 20% likely to be a non-producer, and 57% likely to be a net seller.

We now turn focus to the decisions of the net sellers in Table 8. Simulations show that when no purchasing enterprise is active in the village, the probability of a producer being a net seller is around 60%. If, instead, an informal trader is active in the area, the probability of being a net seller is 67%. When either a formal private trader or the KCC is active that probability increases to around 72%, and when a non-KCC milk cooperative is active the probability of a given producer being a net seller increases to around 78%. Expected volume among net sellers is highest in villages where a formal private trader is active (1316 liters/year), but also quite high where either a non-KCC cooperative or informal private trader is active (1245 and 1230 liters/year respectively). Interestingly, where the KCC is active, although households are more likely to produce and be a net

Table 8: Average[a] Household Production and Net Sales Probability and Expectation

State of Nature	Probability of Producing	Probability of net seller, given production	Expected net sales, given net seller (liters)	"Unconditional"[b] expected net sales (liters)
Marketing Agent in Village				
Milk Co-op	0.683	0.775	1245	659
KCC	0.716	0.725	1065	553
Private Trader	0.797	0.713	1316	748
Informal Trader	0.707	0.670	1230	583
All	0.899	0.942	2480	2100
None	0.649	0.603	953	373

Source: Tegemeo household surveys 2000, 2004, 2007, and author's calculations.
Notes: (a) All explanatory variables not described in the "State of Nature" are held constant at the national mean values. (b) "Unconditional" implies results are not conditioned on being a producer or net seller (all probabilities are conditional on explanatory variables). That is, this column shows the expected net sales of any given household, versus the expected net sales of a given net seller.

seller, expected volume among net sellers is not much higher than if there were no purchasing enterprise in the village (1065 liters/year compared to 953 liters/year). Together, these effects mean the unconditional expected volume of net sales changes substantially depending on whether a purchasing enterprise is present. Where there are none the unconditional expected net sales are 373 liters/year. When the KCC is present the unconditional net sales are expected to be 180 liters higher at 553 liters/year. Rather than the KCC, when an informal private trader is present unconditional expectations are even greater, reaching 583 liters/year, or 659 liters/year if there is a non-KCC cooperative. The highest unconditional expected net sales are in areas where a formal private trader is present, in which case the per-household volume is expected to be nearly 750 liters annually. In other words, as compared to a village where there is no active purchasing enterprise (which described 38% of the sample in 2004 and 41% in 2007), per household net sales are 48% higher when the KCC is actively present, 56% higher when there is an informal purchasing enterprise, 77% higher when a non-KCC cooperative is in the area, and 101% higher when a formal private enterprise is active in the village.

We can also calculate the impact of having more than one type of purchasing agent present. The last row in Table 8 indicates that if all 4 types are present the unconditional expected net sales volume is 2100 liters/year, or 463% greater than the benchmark of no agents.[15] Moreover, due to the non-linearity of the calculations, as shown in equation (5.28), the model allows for interaction effects. That is, notice the sum of the individual effects is 282%, meaning the interaction of having multiple purchasing enterprises present increased expected net sales an additional 181% versus the benchmark. Table D2 in Appendix D includes all combinations of purchasing enterprise types and calculates their interactions.

These results are quite relevant, given the GoK's recent focus on the revival of the KCC in order to promote dairy marketing. It seems that while the KCC's activity does promote dairy farming, this objective might be better achieved by providing an enabling environment for formal private enterprises and farmer-owned cooperatives. Also, rather than loosely tolerating illegal informal private traders or "hawkers," the policy environment should encourage such marketing channels, while of course promoting the continued safety practice of boiling raw milk.

Finally, notice the results indicate that prices have very little significant effect on the decisions of farmers. Higher milk price does significantly increase the probability of a given producer selling milk, as one would expect, but otherwise has no significant impact on behavior. Higher grain prices seem to promote dairy production, which seems counterintuitive since grain is an input for dairy. This may be explained by the fact that

[15] Only approximately 3% of the sample are in a village where all 4 types of enterprise is active, however about 1/3 of households are in a village where at least 2 types are present, underscoring the relevance of interaction effects.

households in areas where rainfall is less reliable (i.e. grain is in lower supply with higher prices) depend on the relatively more reliable income from dairy farming. Other than this, however, prices of inputs, outputs, and substitutes have no significant effects. There are two possible explanations for this outcome. First, recall that time-variant and fixed effects are controlled for at the zone-level. It may be that prices do not vary much within a zone, so the price effects are being captured here.[16] The second possibility is that price elasticities are low because of other rigidities in the market, such as access to purchasing enterprises, credit, electricity, and so on. The latter explanation is supported by anecdotal evidence from the aforementioned allafrica.com article which describes Kenyan farmer's failure to respond to rising world prices with increased production.

[16] This is explored further in Appendix B.

8. CONCLUSION

This study examines the determinants of smallholder household participation in dairy production and marketing. Dairy production holds much income generating potential for rural farmers in Kenya, who hold 85% of the dairy cattle in East Africa (3 million cows). Despite this potential and a history of growth in the sector, recent output levels have not kept pace with increasing domestic and export demand. The objectives of this study are to develop a modeling framework to determine the factors enabling smallholder participation in Kenya's dairy sector, and use the findings of the model to identify strategies to improve smallholder dairy productivity and promote successful commercialization of dairy production.

Traditional market participation models are not sufficient for addressing these objectives, primarily because each uses a double hurdle model hinging on the implicit assumption that all observations are producers. Roughly 1/3 of rural Kenyan households, however, do not produce milk in a given year. Past market participation studies have circumvented this problem by focusing on a sub-population of only producers. While this resolves the problem of data limitation, findings from such analyses cannot be used to estimate the effects on the more general population (i.e. they ignore the decision to produce or not produce milk).

This study therefore develops a new "triple hurdle" model, which allows for the fact that there are two types of market non-participants (non-producers and autarkic producers). Rather than focusing on a sub-population of producers, the triple hurdle allows us to extrapolate implications to the entire population by estimating the expected sales and purchase volumes without condition on being a producer. Results are also used to predict the probabilities of a household being either a non-producer, autarkic producer, net buyer, or net seller separately.

Results indicate there is unexploited potential for smallholder income generation in the dairy market. First, it seems that farm households are more likely to engage in dairy production and marketing in areas where rainfall (and thus crop incomes) are less reliable. Technical education is also an important determinant at every stage of the decision process from production to sales volume among net sellers, which could provide a policy lever for raising national production. Among producers, the use of improved technologies such as grade cows and zero-grazing feeding notably increase the probability of being a net seller and having higher net sales volume, with all coefficients significant at the 1% level in the latter stages of the model.

An enabling environment for producers is an important factor in promoting income generation through dairy. First, because of the productivity gains of artificial insemination and the benefits of cold storage for milk, proximity to electrical infrastructure is a significant determinant of whether producers become net sellers and of how much they sell. Second, due to the substantial investment required to be a dairy farmer, access to credit is a significant factor at every stage. Simulation analysis shows that in a relatively enabling environment, where credit is readily accessible and electricity within 1 km, the unconditional per-household expected net sales are 53% higher than that of the current data mean.

Access to a purchasing enterprise (i.e. selling outside of a household-to-household transaction) is also an important factor at every stage. In a given year about 40% of the sample was in a village where no such enterprise was active. In an area where all purchasing enterprise variables are 0 and all other controls are held at the data means, the unconditional expected net sales for a given household is 373 liters annually. If the KCC is active in the area expectated net sales increase 48%. If, instead, only an informal or only a private retailer is active in the area the expected net sales for any given household are 56% and 101% higher respectively, compared to the benchmark of 274 liters with no purchasing enterprises. Finally, if a non-KCC milk cooperative is active in the village, expected net sales are 77% higher than if only household-to-household transactions were possible. There is also evidence that the interaction of having multiple purchasing agents present can increase unconditional expected sales up to an additional 181% versus the benchmark. These figures are quite relevant to an ongoing debate over Kenyan policy, which currently limits the number of licenses for private dairy purchasing enterprises, and is more focused on reviving the KCC.

Controlling for other factors, evidence of Kenyan farmers' price responsiveness vis-à-vis the dairy market is inconclusive. The lack of significant results may be confirmation of anecdotal evidence that rising world prices have not induced a comparable increase in production that many anticipated. It may also be, however, that price effects, measured using district-level prices, are being confounded as the model controls for zone-level time-varying and fixed effects. Either way, when controlling for prices, several other factors appear highly significant, which highlights the importance of the other constraints controlled for in the model.

Altogether, results from this study indicate that there is indeed unexploited income generation potential for rural smallholders in Kenya's dairy market. With all else equal, however, potential alone will not be enough incentive for farmers to become dairy producers, or allow them to produce in surplus. Rather, tapping that potential will require providing a more enabling environment for producers. This includes continued and increased investment in the electrification of rural areas, technical training extension services and technical colleges, as well as encouraging rural credit markets.

Given their prominent role in the decisions of producers, providing an enabling environment for purchasing enterprises will also encourage increased production. The ongoing revival of the KCC will likely induce increases in production, but results indicate that less centrally controlled, farmer-owned cooperatives and formal private processors are ultimately more effective. Despite the potential for hidden action problems, informal private dairy traders have also proven to be an important part of the rural markets, providing another policy lever to induce economic growth. Such traders should be legitimized, not just tolerated as they are under current policy. An expansive licensing system for so-called "hawkers" could promote development in rural markets and lead to increased production and income generation for farmers while maintaining the safety of consumers.

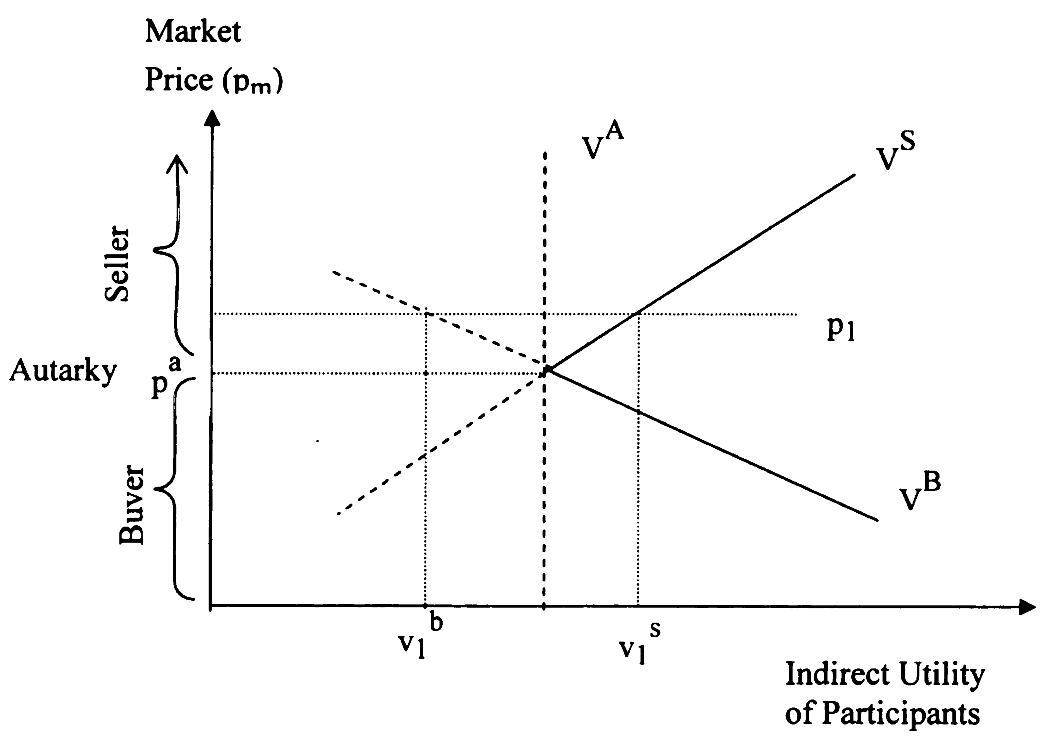
The dairy market will undoubtedly continue to play a substantial role in Kenya's rural economy, and, if properly fostered, could ultimately be an important catalyst for the elusive growth of the country's overall economy.

APPENDIX A: AUTARKY PRICES AND TRANSACTION COSTS

Once a household has determined it will be a producer, the next logical question is why such an agent would rationally choose not to participate in the market. That is, why would this household remain autarkic rather than becoming a buyer or a seller. Traditionally, this behavior has been explained through transactions costs faced by the household.

Borrowing from Key et. al., this point can be illustrated by considering the indirect utility function of a producing household that could be either a buyer or a sellers. First consider the case where there are no transactions costs, or where the price received as a seller is

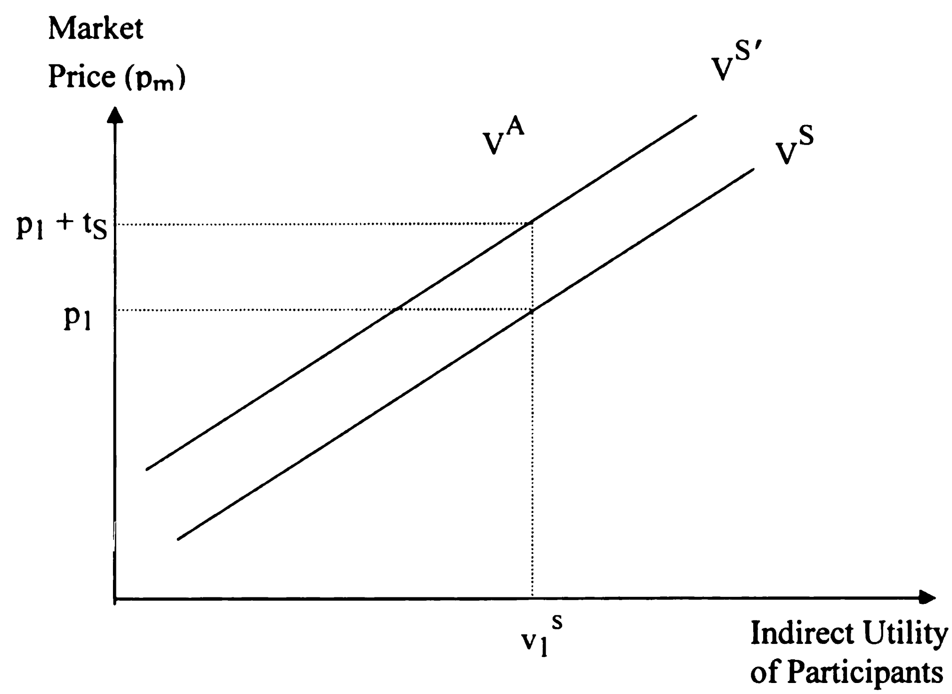
Figure A1. Indirect Utility of Market Participants



the same as that paid as a buyer, and both are identical to the market price. Standard utility maximization theory stipulates that for any given market price, each participant will derive a specific level of utility (the indirect utility). Moreover, all else equal the indirect utility is increasing in price for sellers, and decreasing in price for buyers. Figure 3 demonstrates this relationship for a representative household in a linear example where V^S is the indirect utility as a seller and V^B is that as a buyer. As such, this household, facing a given market price, will have the choice to be either a buyer or a seller. In the example illustrated in Figure 3, suppose this household faces market price ($p = p_1$). Then maximizing their utility as a buyer results in their indirect utility level of (v_1^b), while doing so as a seller results in the indirect utility of (v_1^s). Since is plain to see that ($v_1^b < v_1^s$), this household would be a seller. Moreover, there exists a singular price where the household would be indifferent between buying and selling (p^a). This is the price at which households would choose not to participate in the market, or their autarky price.

Now we can show how the autarky price becomes a range of prices when relaxing the assumption of zero transaction costs. Such costs drive a wedge between prices received as a seller and those paid as a buyer, previously assumed to be identical to the market price. Specifically, transaction costs raise the price effectively paid by buyers, and lowers the price effectively received by sellers. For the case of sellers, this is illustrated in Figure 4, as the shift from V^S to $V^{S'}$ after the introduction of transaction costs faced by the seller (t_s). Since the seller effectively receives the market price, less the transaction costs they face ($p_m - t_s$), the market price must be higher for them to achieve the same utility as if there were no transaction costs. In other words, under no transaction costs, at market price p_1, the seller receives utility v_1^s. In the presence of transaction costs, however, in order to achieve the same utility, the seller must face a market price greater than p_1 by a margin equal to said cost $(p_1 + t_s)$.[17]

Figure A2. Indirect Utility of Sellers Under Transaction Costs

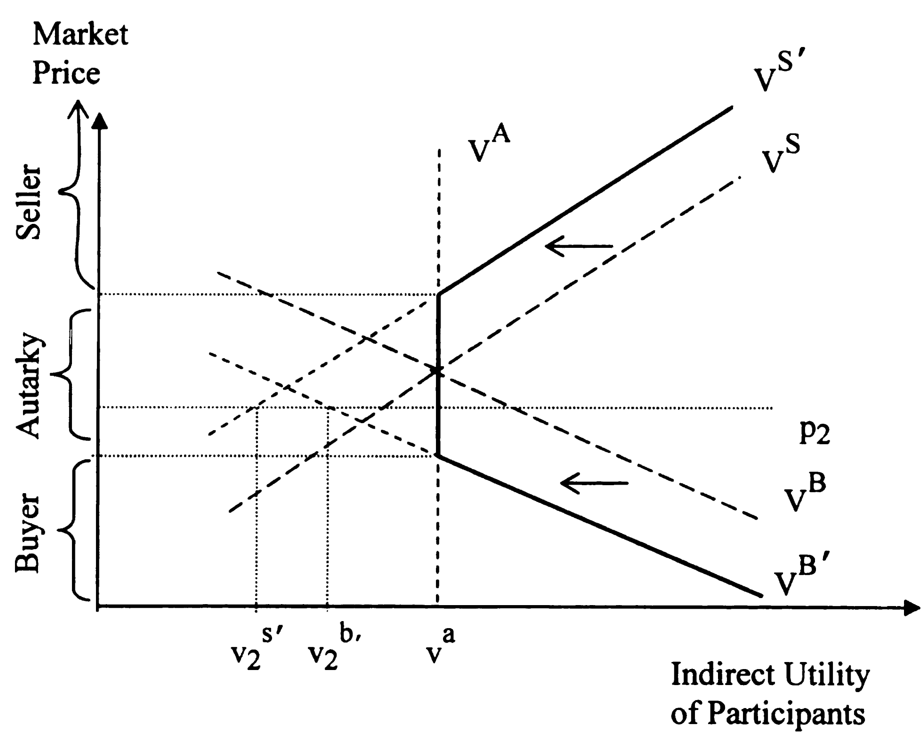


Similarly, it can be shown that transaction costs faced by a buyer (t_b) will shift the buyers indirect utility function to the left from V^B to $V^{B'}$, as seen in Figure 5. On the other hand, the indirect utility of a household not participating in the market will not be affected by transaction costs whatsoever, thus there is no change in the indirect utility of an autarkic household (V^A). The result is that, rather than a singular price at which this representative household would not participate in the market, there is a full range of

[17] In this example, transaction costs have been treated as proportional to the quantity of the good being traded. However, the principal is still applicable in the presence of fixed transaction costs.

market prices where this household would be autarkic. For example, if the market price is p_2, as illustrated in Figure 5, then in the presence of transaction costs if this household were to be a buyer they would achieve the utility level $v_2^{b'}$, which is greater than the utility level they would achieve as a seller, $v_2^{b'}$. However, both of these are less than v_a, the utility this household achieves by not participating in the market at all. As such, this household, facing this market price, would rationally behave as an autarkic household. Note that where this price line intersects the original indirect utility functions (V^S and V^B) is irrelevant, since these points don't reflect any feasible state of reality under the specified assumptions.

Figure A3. Indirect Utility of Market Participants Under Transaction Costs



As previously mentioned, these transaction costs have been the focus of the majority of existing market participation studies. While these costs are clearly a key determining factor, the makeup of a households indirect utility also clearly bears weight in their marketing outcomes, yet has traditionally been given much less consideration. Specifically, we will argue that some variables, such as realized output, may not be as clearly defined a priori choice variables as previous studies have implicitly assumed, and lagged investment decisions will affect achievable utility levels.

APPENDIX B: FIRST STAGE ESTIMATION WITH AND WITHOUT PRICES

Table B1: Production Probit Coefficients with and without Controlling for Prices

	(1) Tier 1 With Prices	(2) Tier 1 Without Prices
	(i)	(ii)
zone==3	1.424***	1.079***
	(5.10)	(3.97)
zone==4	1.736***	1.400***
	(4.30)	(3.82)
zone==5	2.164***	1.458***
	(4.19)	(3.43)
zone==6	2.122***	1.577***
	(5.01)	(5.19)
zone==7	2.312***	1.791***
	(4.61)	(4.04)
zone==8	1.616***	1.298***
	(4.53)	(4.34)
zone==9	1.217***	1.121***
	(2.60)	(2.80)
zoneyr==2	-0.715	-0.793
	(1.13)	(1.56)
zoneyr==3	0.174	<dropped>
	(0.48)	
zoneyr==4	-0.036	-0.112
	(0.17)	(0.47)
zoneyr==5	0.057	0.085
	(0.25)	(0.37)
zoneyr==6	-0.002	-0.088
	(0.01)	(0.35)
zoneyr==7	<dropped>	-0.134
		(0.52)
zoneyr==8	-0.008	-0.060
	(0.03)	(0.30)
zoneyr==9	0.761**	0.422
	(2.02)	(1.28)
year = 2007	0.440	-0.001
	(1.60)	(0.01)
population density by division (persons/squared mile)	-0.000	-0.000
	(1.41)	(0.76)
distance from motorable road in km	0.017	0.013
	(0.58)	(0.47)
distance to vet services(kms)	0.014**	0.010
	(1.97)	(1.40)
distance to electricity (kms)	-0.009	-0.009
	(1.45)	(1.48)
household has short rain season	0.006	0.042
	(0.04)	(0.24)
share of household heads without education	-0.605	-1.000**
	(1.42)	(2.45)
milk cooperative in village	0.094	0.065
	(0.97)	(0.67)

Table B1 (continuted)

kenya creamery company in village	0.188*	0.203*
	(1.78)	(1.93)
private processors/traders in village	0.448***	0.385***
	(5.15)	(4.57)
informal trader/hawker in village	0.162**	0.202**
	(2.04)	(2.55)
share hh in village receiving credit	0.306*	0.414***
	(1.95)	(2.75)
lagged acres owned	0.059***	0.058***
	(3.74)	(3.73)
lagged asset value (,000 2007 ksh)	0.002**	0.002**
	(2.16)	(2.12)
past 7 year average main season rainfall (mm)	-0.000	0.000
	(0.49)	(0.12)
past 7 year variance main season rainfall (mm2)	0.000***	0.000***
	(3.42)	(3.25)
did hh experience a prime age death (15 to 59)	-0.124	-0.126
	(0.95)	(0.96)
female headed household	-0.088	-0.082
	(1.04)	(0.97)
1 to 4 years	-0.002	-0.008
	(0.02)	(0.08)
5 to 8 years	0.149	0.127
	(1.40)	(1.20)
9 to 12 years	0.173	0.128
	(1.36)	(1.02)
more than 12 years (or some college)	0.314*	0.259
	(1.88)	(1.56)
age of household head	0.009***	0.007**
	(2.88)	(2.36)
adult equivalents	0.054***	0.058***
	(3.74)	(3.95)
real district price of milk (2007 ksh/litre)	-0.018	
	(1.14)	
real district price of maize (2007 ksh/kg)	0.117***	
	(4.54)	
real district price of tomatoes (2007 ksh/kg)	0.005	
	(0.13)	
price of 50 kg dap bag (2007 ksh)	-0.000	-0.001
	(0.50)	(0.84)
Constant	-3.381*	-0.916
	(1.78)	(0.53)
Observations	2550	2550

z statistics robust to heterokedasticity and autocorrelation in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

One way to mitigate concerns over the unsatisfactory price expectations in the first tier, which assumes perfect foresight for one period, is to examine their overall effect on coefficients in that stage (recall, the likelihood function is separable, so coefficients in latter stages of the model are not affected by the choice of price expectation model). To do this we will compare the results used in the study, found in column (i), from estimates dropping prices altogether (i.e. results assuming complete lack of expectation), found in column (ii). The effects of most interest to the study are highlighted in yellow. Although this is an admittedly subjective assessment, this exercise demonstrates that results are fairly robust to misspecification of price expectations. The coefficient on distance to electricity is virtually unchanged. The coefficient on credit prevalence is significant in both models, though the effect is estimated to be slightly greater in the model without prices. Coefficients on purchasing agents are comparable in either model (i.e. sign is unchanged, with little change in magnitude and z-scores).

Notice, also, that when prices are dropped the zone-level time-varying (*zoneyr*) and fixed (*zone*) effects coefficients change substantially. This lends support to the notion that these effects are highly correlated with prices and therefore price expectations. Table B2 illustrates this further by showing the mean of prices within zones in each year, with the standard deviations in parentheses. Clearly, the variations in prices relative to prevailing prices are quite small within the unit of fixed effects controlled for by the model. Concerns over unsatisfying price expectations, therefore, can be further assuaged by the fact that much of that effect is controlled for by other means.

Table B2: Price Distribution by Zone Over Time

	Year							
	2004				2007			
	Mean Price in 2007 Kenyan Schillings (std. dev.)							
Agro-ecological Zone	Milk (kg)	Maize (kg)	Tomato (kg)	DAP (50 kg)	Milk (kg)	Maize (kg)	Tomato (kg)	DAP (50 kg)
Coastal Lowlands	27	19	13	1828	17	14	7	1800
	(.00)	(.00)	(.50)	(.00)	(.00)	(.26)	(.37)	(.00)
Eastern Lowlands	29	18	13	1778	29	13	8	1793
	(1.09)	(.70)	(.00)	(38.10)	(3.15)	(1.77)	(.00)	(17.30)
Western Lowlands	30	19	11	1896	23	14	11	1811
	(3.69)	(3.73)	(.00)	(.00)	(1.98)	(.55)	(.00)	(12.40)
Western Transitional	27	17	10	1848	19	12	9	1800
	(3.39)	(.69)	(1.41)	(30.84)	(.46)	(.00)	(.00)	(.00)
High Potential Maize Zone	22	16	11	1840	16	11	8	1717
	(3.57)	(1.81)	(.92)	(56.66)	(1.97)	(1.42)	(1.02)	(55.83)
Western Highlands	29	20	8	1896	22	14	9	1800
	(2.66)	(1.48)	(1.36)	(.00)	(2.45)	(.55)	(.15)	(.00)
Central Highlands	20	17	11	1787	15	13	9	1787
	(1.05)	(1.06)	(2.57)	(81.53)	(.85)	(.63)	(.74)	(31.57)
Marginal Rain Shadow	16	18	13	1760	13	12	8	1650
	(.00)	(.00)	(.00)	(.00)	(.00)	(.00)	(.00)	(.00)

Source: Tegemeo household survey data 2004 and 2007. Household-level mean district price with standard deviation in parentheses.

APPENDIX C: FULL REGRESSION RESULTS FROM 3-STAGE ANALYSIS

Table C1: Full Regression Results from 3-stage Analysis

Tier 1: Production Probit						
	Coef.	Std. Err.	Z	P>z	[95% Conf.	Interval]
_Izone_3	1.424008	.2790454	5.10	0.000	.8770893	1.970927
_Izone_4	1.735902	.4037479	4.30	0.000	.9445706	2.527233
_Izone_5	2.163969	.516053	4.19	0.000	1.152524	3.175415
_Izone_6	2.122056	.4233342	5.01	0.000	1.292336	2.951775
_Izone_7	2.312115	.5020647	4.61	0.000	1.328087	3.296144
_Izone_8	1.616275	.3570624	4.53	0.000	.9164452	2.316104
_Izone_9	1.217052	.468911	2.60	0.009	.2980029	2.1361
_Izoneyr_2	-.714738	.6310479	-1.13	0.257	-1.951569	.5220933
_Izoneyr_3	.1736775	.3651059	0.48	0.634	-.5419169	.8892718
_Izoneyr_4	-.0358781	.2083467	-0.17	0.863	-.4442302	.372474
_Izoneyr_5	.0573264	.2321219	0.25	0.805	-.3976243	.512277
_Izoneyr_6	-.0016131	.2361168	-0.01	0.995	-.4643936	.4611673
_Izoneyr_8	-.0077291	.3041176	-0.03	0.980	-.6037886	.5883305
_Izoneyr_9	.7606952	.3767884	2.02	0.043	.0222035	1.499187
year07	.4396577	.2751114	1.60	0.110	-.0995507	.9788662
popden	-.0002197	.0001557	-1.41	0.158	-.0005249	.0000854
dmtroad	.0169623	.0290633	0.58	0.559	-.0400007	.0739254
dvet	.0139351	.0070782	1.97	0.049	.0000622	.0278081
delect	-.0088163	.0060737	-1.45	0.147	-.0207206	.0030879
shrtseas	.0061699	.1728592	0.04	0.972	-.3326279	.3449677
ed0div	-.6053069	.4276809	-1.42	0.157	-1.443546	.2329324
milkcoop	.094362	.0973655	0.97	0.332	-.0964709	.2851949
kcc	.1884748	.1060639	1.78	0.076	-.0194067	.3963563
privtrad	.4478607	.0869933	5.15	0.000	.2773569	.6183645
inftrade	.1624473	.0796828	2.04	0.041	.0062719	.3186228
credprev	.3056186	.156901	1.95	0.051	-.0019016	.6131389
acrown_1	.059243	.0158554	3.74	0.000	.028167	.0903191
raval_1	.0015625	.0007234	2.16	0.031	.0001447	.0029803
avgrain7	-.0003614	.0007382	-0.49	0.624	-.0018083	.0010855
varrain7	4.13e-06	1.21e-06	3.42	0.001	1.76e-06	6.49e-06
pad	-.1243207	.1314604	-0.95	0.344	-.3819783	.133337
femhed	-.0879407	.0845604	-1.04	0.298	-.253676	.0777947
ed1_4	-.0021821	.1024888	-0.02	0.983	-.2030565	.1986923
ed5_8	.149312	.1063503	1.40	0.160	-.0591307	.3577547
ed9_12	.1726056	.1265008	1.36	0.172	-.0753314	.4205426
edplus	.3138771	.1670585	1.88	0.060	-.0135515	.6413057
hedage	.0087469	.0030419	2.88	0.004	.0027848	.014709
ae	.0543839	.0145385	3.74	0.000	.0258889	.0828788
rmilkp	-.0183533	.0160878	-1.14	0.254	-.0498849	.0131783
rcornp	.1167339	.0257283	4.54	0.000	.0663072	.1671605
rtomatp	.0045532	.0358441	0.13	0.899	-.0657	.0748064
ractdap	-.0004858	.0009703	-0.50	0.617	-.0023874	.0014159
_cons	-3.380791	1.904467	-1.78	0.076	-7.113478	.3518949

Table C1 (continued)

Tier 2: Market Participation Ordered Probit						
_Izone_3	-.2464386	.4535206	-0.54	0.587	-1.135323	.6424453
_Izone_4	.0413585	.5703074	0.07	0.942	-1.076423	1.15914
_Izone_5	.9798834	.7106406	1.38	0.168	-.4129466	2.372713
_Izone_6	.5516018	.6483378	0.85	0.395	-.7191169	1.822321
_Izone_7	.2384912	.6600936	0.36	0.718	-1.055269	1.532251
_Izone_8	-.2398228	.6136676	-0.39	0.696	-1.442589	.9629437
_Izone_9	1.084873	.7435123	1.46	0.145	-.3723844	2.54213
_Izoneyr_2	.2259821	.8226565	0.27	0.784	-1.386395	1.838359
_Izoneyr_3	-.238402	.5647683	-0.42	0.673	-1.345328	.8685235
_Izoneyr_4	.3038556	.5351889	0.57	0.570	-.7450953	1.352806
_Izoneyr_5	.0272478	.5288153	0.05	0.959	-1.009211	1.063707
_Izoneyr_6	.2871383	.5064966	0.57	0.571	-.7055767	1.279853
_Izoneyr_7	.2429013	.5635988	0.43	0.666	-.8617321	1.347535
_Izoneyr_8	.2500005	.5108929	0.49	0.625	-.7513311	1.251332
year07	-.291669	.5967651	-0.49	0.625	-1.461307	.8779691
popden	-.00006	.0001554	-0.39	0.700	-.0003646	.0002446
dmtroad	.0564862	.0377779	1.50	0.135	-.0175572	.1305296
dvet	.0041202	.0071513	0.58	0.565	-.0098962	.0181366
delect	-.0201802	.0059865	-3.37	0.001	-.0319136	-.0084469
shrtseas	-.1733924	.1876567	-0.92	0.355	-.5411927	.194408
ed0div	.1011895	.5759857	0.18	0.861	-1.027722	1.230101
milkcoop	.493414	.123426	4.00	0.000	.2515034	.7353246
kcc	.3363483	.1281078	2.63	0.009	.0852616	.587435
privtrad	.3007941	.1023294	2.94	0.003	.1002322	.501356
inftrade	.1795128	.0986933	1.82	0.069	-.0139226	.3729481
credprev	.5405856	.1841465	2.94	0.003	.1796651	.9015061
acrown_1	.005976	.012622	0.47	0.636	-.0187626	.0307146
raval_1	.0003899	.0002207	1.77	0.077	-.0000427	.0008224
avgrain7	-.0015172	.0009402	-1.61	0.107	-.0033599	.0003254
varrain7	2.88e-06	1.55e-06	1.86	0.063	-1.51e-07	5.91e-06
posrain	.0000587	.0005201	0.11	0.910	-.0009608	.0010782
negrain	.0006812	.0018114	0.38	0.707	-.002869	.0042314
pad	.0498296	.1493622	0.33	0.739	-.242915	.3425742
chrnill	.0488312	.0668476	0.73	0.465	-.0821878	.1798501
femhed	.1825046	.0898829	2.03	0.042	.0063374	.3586719
ed1_4	.1762784	.1112448	1.58	0.113	-.0417574	.3943142
ed5_8	.1417027	.1075655	1.32	0.188	-.0691218	.3525273
ed9_12	.1982075	.1403715	1.41	0.158	-.0769155	.4733306
edplus	.6326099	.1755092	3.60	0.000	.2886181	.9766016
hedage	-.0048684	.0032836	-1.48	0.138	-.0113041	.0015673
ae	-.0183661	.0147174	-1.25	0.212	-.0472117	.0104795
ngrdc_1	.1180563	.0393565	3.00	0.003	.0409189	.1951937
zgraze_1	.458023	.1327833	3.45	0.001	.1977725	.7182735
acresf	.037266	.0148255	2.51	0.012	.0082084	.0663235
rmilkp	.0357234	.0183499	1.95	0.052	-.0002418	.0716887
rcornp	.0154896	.031099	0.50	0.618	-.0454633	.0764425
rtomatp	-.0555735	.0462866	-1.20	0.230	-.1462936	.0351465
ractdap	-.0008929	.0011445	-0.78	0.435	-.003136	.0013502
a1	-2.147293	2.50341	-0.86	0.391	-7.053886	2.759299
a2	-1.672629	2.50286	-0.67	0.504	-6.578144	3.232887

Table C1 (continued)

Tier 3: Net Purchases Lognormal						
Izone_3	-1.427801	.8023134	-1.78	0.075	-3.000306	.1447044
Izone_4	-2.558261	1.20205	-2.13	0.033	-4.914235	-.2022868
Izone_5	-2.553028	1.640041	-1.56	0.120	-5.767449	.661392
Izone_6	-1.754372	1.5002	-1.17	0.242	-4.69471	1.185966
Izone_7	-3.011166	1.386517	-2.17	0.030	-5.728689	-.2936432
Izone_8	-1.685872	1.408936	-1.20	0.231	-4.447336	1.075592
Izone_9	-.9097269	1.558529	-0.58	0.559	-3.964388	2.144934
Izoneyr_2	-.6508669	1.518806	-0.43	0.668	-3.627671	2.325937
Izoneyr_3	.1386056	1.248944	0.11	0.912	-2.309281	2.586492
Izoneyr_5	.8871922	.776995	1.14	0.254	-.63569	2.410074
Izoneyr_6	.1075194	.6433799	0.17	0.867	-1.153482	1.368521
Izoneyr_7	.8165686	.565975	1.44	0.149	-.2927219	1.925859
Izoneyr_8	1.487881	.8807469	1.69	0.091	-.2383511	3.214113
Izoneyr_9	-.5599216	.8927159	-0.63	0.531	-2.309613	1.189769
year07	-.7121697	.8332874	-0.85	0.393	-2.345383	.9210435
popden	.0004191	.0002956	1.42	0.156	-.0001602	.0009985
dmtroad	.0326371	.0798299	0.41	0.683	-.1238267	.1891009
dvet	-.0049592	.0233012	-0.21	0.831	-.0506287	.0407103
delect	-.0159563	.0144903	-1.10	0.271	-.0443567	.0124442
shrtseas	.5475004	.4736763	1.16	0.248	-.380888	1.475889
ed0div	-1.287358	1.360274	-0.95	0.344	-3.953445	1.37873
milkcoop	-.687645	.3085426	-2.23	0.026	-1.292377	-.0829126
kcc	.5255982	.3205409	1.64	0.101	-.1026504	1.153847
privtrad	-.0108362	.1987742	-0.05	0.957	-.4004264	.378754
inftrade	-.2366261	.2068219	-1.14	0.253	-.6419896	.1687374
credprev	-.6118422	.5096947	-1.20	0.230	-1.610825	.3871411
acrown_1	.0019603	.0273054	0.07	0.943	-.0515572	.0554779
raval_1	.0006249	.0003359	1.86	0.063	-.0000335	.0012833
avgrain7	.0011342	.0020649	0.55	0.583	-.002913	.0051814
varrain7	-.0000111	7.09e-06	-1.56	0.119	-.000025	2.84e-06
posrain	-.0017596	.0012991	-1.35	0.176	-.0043059	.0007866
negrain	-.0016624	.0040639	-0.41	0.682	-.0096274	.0063027
pad	.0118139	.3763805	0.03	0.975	-.7258783	.7495062
chrnill	-.2531904	.1909976	-1.33	0.185	-.6275388	.1211581
femhed	.1364443	.2434559	0.56	0.575	-.3407205	.6136091
ed1_4	.0089356	.3077968	0.03	0.977	-.5943349	.6122062
ed5_8	.0801894	.2835685	0.28	0.777	-.4755948	.6359735
ed9_12	.3018563	.3017566	1.00	0.317	-.2895757	.8932884
edplus	.5109987	.437426	1.17	0.243	-.3463406	1.368338
hedage	.0079499	.0078121	1.02	0.309	-.0073615	.0232613
ae	.0589357	.0323738	1.82	0.069	-.0045158	.1223873
ngrdc_1	.0337963	.1064726	0.32	0.751	-.1748863	.2424788
zgraze_1	.0972596	.3077497	0.32	0.752	-.5059187	.7004379
acresf	.0341211	.0236615	1.44	0.149	-.0122546	.0804968
rmilkp	-.0125175	.0378708	-0.33	0.741	-.0867428	.0617078
rcornp	-.0536703	.0775881	-0.69	0.489	-.2057401	.0983995
rtomatp	-.1703158	.1132577	-1.50	0.133	-.3922968	.0516652
ractdap	-.0033099	.0029027	-1.14	0.254	-.0089991	.0023793
_cons	14.41093	6.424981	2.24	0.025	1.818203	27.00367
sigma3	1.26599	.0636911	19.88	0.000	1.141158	1.390822

Table C1 (continued)

Tier 3: Net Sales Lognormal						
Izone_3	-1.79998	.5818396	-3.09	0.002	-2.940364	-.6595948
Izone_4	-1.128672	.682936	-1.65	0.098	-2.467202	.2098579
Izone_5	-2.187541	.7752991	-2.82	0.005	-3.7071	-.6679832
Izone_6	-1.34774	.7234823	-1.86	0.062	-2.765739	.0702593
Izone_7	-2.165497	.7228024	-3.00	0.003	-3.582164	-.7488303
Izone_8	-2.060374	.6790624	-3.03	0.002	-3.391312	-.7294361
Izone_9	-.2687196	.7305747	-0.37	0.713	-1.70062	1.163181
Izoneyr_2	.5878002	.9695167	0.61	0.544	-1.312417	2.488018
Izoneyr_3	1.425091	.5187213	2.75	0.006	.4084162	2.441766
Izoneyr_5	1.368268	.3549289	3.86	0.000	.6726204	2.063916
Izoneyr_6	1.158611	.3723851	3.11	0.002	.4287496	1.888472
Izoneyr_7	1.121634	.3534382	3.17	0.002	.4289076	1.81436
Izoneyr_8	1.289079	.3941249	3.27	0.001	.5166083	2.061549
Izoneyr_9	.6601954	.4304053	1.53	0.125	-.1833835	1.503774
year07	-1.064519	.4019184	-2.65	0.008	-1.852265	-.2767735
popden	.0003783	.0002566	1.47	0.140	-.0001246	.0008813
dmtroad	-.0467295	.0333041	-1.40	0.161	-.1120043	.0185454
dvet	.0001172	.0085898	0.01	0.989	-.0167186	.016953
delect	-.0122519	.006985	-1.75	0.079	-.0259424	.0014385
shrtseas	.2005916	.2063152	0.97	0.331	-.2037787	.6049619
ed0div	.1449902	.5637022	0.26	0.797	-.9598459	1.249826
milkcoop	.2675182	.1020291	2.62	0.009	.0675447	.4674916
kcc	.1110111	.1055533	1.05	0.293	-.0958696	.3178918
privtrad	.3225243	.0921045	3.50	0.000	.1420028	.5030458
inftrade	.2554078	.0919739	2.78	0.005	.0751423	.4356733
credprev	.328296	.1602627	2.05	0.041	.0141868	.6424052
acrown_1	.0187362	.0173038	1.08	0.279	-.0151786	.052651
raval_1	.0000145	.000077	0.19	0.850	-.0001364	.0001655
avgrain7	.0006723	.0010021	0.67	0.502	-.0012918	.0026364
varrain7	2.32e-06	1.39e-06	1.67	0.096	-4.10e-07	5.04e-06
posrain	.0007703	.0004725	1.63	0.103	-.0001558	.0016964
negrain	.0017614	.0015046	1.17	0.242	-.0011875	.0047103
pad	-.0530532	.1697255	-0.31	0.755	-.385709	.2796027
chrnill	-.119718	.0736187	-1.63	0.104	-.264008	.0245719
femhed	.0205735	.0944823	0.22	0.828	-.1646084	.2057554
ed1_4	-.1143323	.1126027	-1.02	0.310	-.3350296	.106365
ed5_8	-.053189	.115315	-0.46	0.645	-.2792024	.1728243
ed9_12	.219704	.1340457	1.64	0.101	-.0430208	.4824288
edplus	.2893219	.1465252	1.97	0.048	.0021379	.576506
hedage	.0043911	.0031912	1.38	0.169	-.0018635	.0106456
ae	.0113534	.0135985	0.83	0.404	-.0152992	.0380061
ngrdc_1	.0962648	.0279376	3.45	0.001	.0415082	.1510215
zgraze_1	.2035985	.1053343	1.93	0.053	-.0028529	.41005
acresf	.0355312	.0154783	2.30	0.022	.0051943	.0658682
rmilkp	-.008844	.0192089	-0.46	0.645	-.0464929	.0288048
rcornp	-.0155711	.0383413	-0.41	0.685	-.0907186	.0595764
rtomatp	-.0035829	.043068	-0.08	0.934	-.0879946	.0808287
ractdap	-.000688	.0009654	-0.71	0.476	-.0025802	.0012043
_cons	7.919367	2.457666	3.22	0.001	3.10243	12.7363
sigma4	1.104533	.0320134	34.50	0.000	1.041788	1.167278

APPENDIX D: COMPLETE SIMULATION ANALYSIS

Table D1: Average[a] Household Production and Participation Probability and Expectation by Credit Prevalence and Distance to Veterinary Services

	Distance to Electricity (km)			
Credit Prevalence (% of households in division receiving)	1	2	4	10
Probability of a given household being a producer				
95	0.80	0.79	0.79	0.77
75	0.78	0.78	0.77	0.75
50	0.76	0.75	0.75	0.73
25	0.73	0.73	0.72	0.70
Probability of a given producer being a net seller				
95	0.83	0.82	0.81	0.78
75	0.80	0.79	0.78	0.74
50	0.76	0.75	0.73	0.70
25	0.71	0.71	0.69	0.65
Expected net sales volume for a given net seller (liters)				
95	1566	1547	1501	1403
75	1467	1449	1405	1314
50	1351	1335	1295	1210
25	1245	1230	1193	1115
Unconditional[b] expected net sales volume (liters)				
95	1030	1008	954	841
75	910	889	839	735
50	772	753	708	615
25	648	631	591	508

Note: (a) All explanatory variables not shown here are held constant at the national mean values.
(b) "Unconditional" implies results are not conditioned on being a producer or net seller (all probabilities are conditional on explanatory variables). That is, these rows show the expected net sales of any given household, versus the expected net sales of a given net seller

Table D2: Average[a] Household Production and Net Sales Probability and Expectation with Multiple Enterprise Interactions

Active agents in village	Probability of Producing	Probability of net seller, given producer	Expected net sales, given net seller (liters)	Unconditional[b] expected net sales (liters)	Total Effect (% Over Benchmark)	Interaction Effect[c] (% of Benchmark)
Benchmark Case						
None	0.649	0.603	953	373	-	-
One Type of Enterprise						
Milk Co-op	0.683	0.775	1245	659	0.77	-
KCC	0.716	0.725	1065	553	0.48	-
Private Trader	0.797	0.713	1316	748	1.00	-
Informal Trader	0.707	0.670	1230	583	0.56	-
Two Types of Enterprise						
KCC & Milk Co-op	0.747	0.862	1391	897	1.40	0.15
KCC & Private Trader	0.846	0.816	1470	1014	1.72	0.23
KCC & Informal Trader	0.768	0.782	1375	825	1.21	0.17
Milk Co-op & Private Trader	0.822	0.854	1719	1208	2.24	0.47
Milk Co-op & Informal Trader	0.840	0.771	1698	1099	1.95	0.62
Private & Informal Traders	0.739	0.825	1608	980	1.63	0.06
Three Types of Enterprise						
KCC & Co-op & Private Trader	0.867	0.918	1921	1529	3.10	0.85
KCC & Co-op & Informal Trader	0.796	0.898	1796	1284	2.44	0.63
KCC & Priv & Informal Traders	0.881	0.860	1898	1438	2.85	0.80
Co-op & Private & Informal Traders	0.862	0.892	2219	1705	3.57	1.24
Four Types of Enterprise						
All Types	0.899	0.942	2480	2100	4.63	1.81

Source: Tegemeo household surveys 2000, 2004, 2007, and author's calculations.

Notes: (a) All explanatory variables not described in the "State of Nature" are held constant at the national mean values. (b) "Unconditional" implies results are not conditioned on being a producer or net seller (all probabilities are conditional on explanatory variables). That is, this column shows the expected net sales of any given household, versus the expected net sales of a given net seller. (c) Interaction effect is the total effect minus the sum of individual effects.

APPENDIX E: MODEL EVALUATION

As always, it is prudent to evaluate the performance of the model as a whole, beyond the significance of the individual coefficients it estimates. One way to evaluate the probabilistic portion of the triple-hurdle would be to compare the average predicted probability for every possible outcome with the actual share of the sample where that outcome is observed. These results are in Table E1, and suggest that the model is fairly accurate overall.

Table E1: Mean Predicted Probability and Sample Share

Possible Outcomes:	Average Predicted Probability	Actual Share of Sample
Non-Producer	.313	.313
Producer		
Net Buyer	.116	.114
Autarkic	.081	.080
Net Seller	.490	.493
Total	1.00	1.00

Source: Tegemeo household survey data and author's calculations.

While Table E1 demonstrates the performance of the model in aggregate, the evaluation does not inform us as to the model's performance with respect to individual observations. To that end, Table E2 presents a cross tabulation of "most likely" predicted outcomes and actual outcomes for each observation. "Most likely" is defined as the highest predicted probability. For example, if an observations predicted probability of being a non-producer is .35, a net buyer is .25, autarkic is .1, and net seller is .3, we would say they are most likely to be a non-producer. Once again, based on this evaluation the model is doing a fairly good job of predicting households' role in the market, with nearly 2/3 of the sample's most likely outcome being realized.

Table E2: Most Likely Outcome by Actual Observed Outcome

Actual Value	Most Likely Outcome[a] Non-producer	Net Buyer	Autarky	Net Seller	Total
Non-producer	**494**	13	0	290	797
Net Buyer	112	**24**	0	155	291
Autarky	67	12	**0**	126	205
Net Seller	115	13	0	**1,089**	1,257
Total	828	62	0	1,660	2,550

Source: Tegemeo household survey data and author's calculations. Note (a) 63% "Correctly" predicted.

Of course, none of the evaluations above consider results from the third stage of the triple-hurdle model. A traditional R-squared would be inappropriate, since the predicted expected values are highly non-linear. In other non-linear models (e.g. log-log equations) the squared correlation between predicted and actual values is considered comparable to the R-squared from a linear regression. The analogous measure for the triple-hurdle would be the squared correlation of the unconditional expected value and actual value. For net sales, that squared correlation is 0.105 (correlation is .3235).

REFERENCES

AfricaNews.com (2007). Kenya: Milk Business Growing Fast. Retrieved on November 6, 2008 from: http://www.africanews.com/site/list_messages/14090

AllAfrica.com (2008). Country Faces Shortage of Milk for Export. Retrieved on November 6, 2008 from: http://allafrica.com/stories/printable/200809220332.html

Alderman, Harold, Behrman R.Jere, Hans-Peter Kohler, Maluccio A. John., and Watkins C.Susan, 2001. "Attrition in longitudinal household survey data: Some tests for three developing country samples". *Demographic Research 5:78-124.*

Barrett, C., and B. Swallow. 2006. Fractal Poverty Taps. *World Development* 34.1: 1-15.

Bellemare, M.F., and Barrett, C.B. 2006. An Ordered Tobit Model of Market Participation: Evidence from Kenya and Ethiopia. *American Journal of Agricultural Economics* 88.2: 324-337.

Burke, W.J., T.S. Jayne, H.A. Freeman and P. Kristjansen, 2007. "Factors Associated with Rural Farm Households' Movements into and out of Poverty: The Rising Importance of Livestock" International Development Working Paper 90, Department of Agricultural Economics, Michigan State University.

Carter, M., and C. Barrett. 2006. The Economics of Poverty Traps and Persistent Poverty: An Asset-based Approach. *Journal of Development Studies* 42.2: 178-199.

Ellis, F. 2007. Strategic Dimensions of Rural Poverty Reduction in Sub-Saharan Africa. Paper presented at the workshop: Rural Development Restrospect and Prospect: A Workshop for Judith Heyer. Oxford, September 14-15, 2007.

Goetz, S.J. 1992. A Selectivity Model of Household Food Marketing Behavior in Sub-Saharan Africa. *American Journal of Agricultural Economics* 74.2: 444-452.

Hazell, P. and S. Haggblade, 1993. "Farm-Nonfarm Grwotn Linkages and the Welfare of the Poor" in Lipton, M. and J. van der Gaag (eds), Including the Poor, Proceedings of a Symmposium Organized by the World Band and the International Food Policy Research Institute, Washington D.C.: World Bank: 190 - 204.

Heckman, J.J. 1976. The Common Structure of Statistical Models of Truncation, Sample Selection, and Limited Dependant Variables and a Simple Estimator for Such Models. *Annals of Economic and Social Measurement* 5: 475-492.

Hoddinott, J. 2006. Shocks and their Consequences Across and Within Households in Rural Zimbabwe. *Journal of Development Studies* 42.2: 301-321.

Holloway, G.J., C.B. Barrett, and S.K. Ehui. 2001 "The Double Hurdle Model in the Presence of Fixed Costs" Applied Economics and Management Working Paper, Cornell University.

Jayne, T.S., T. Yamano, M.T. Weber, D. Tschirley, R. Benfica, A. Chapoto, and B. Zulu. 2003. Smallholder Income and Land Distribution in Africa: Implications for Poverty Reduction Strategies. *Food Policy* 28: 253-275.

Johnston, B.F. and J.W. Mellor, 1961. "The Role of Agriculture in Economic Development", *American Economic Review* 51 (4): 566 - 93

Karanja, A.M. 2003 "The Dairy Industry in Kenya: The Post-Liberalization Agenda." Working Paper 1, Tegemeo Institute, Nairobi, Kenya.

Key, N., E. Sadoulet, and A. DeJanvry. 2000. Transactions Costs and Agricultural Household Supply Response. *American Journal of Agricultural Economics* 82.2: 245-259.

Kijima, Y. T. Yamano, and I. Baltenweck. 2006. "Emerging Markets after Liberalization: Evidence from the Raw Milk Market in Rural Kenya." FASID Discussion Paper (*series forthcoming*).

Krishna, A. 2004. Escaping Poverty and Becoming Poor: Who Gains, Who Loses, and Why? *World Development* 32.1: 121-136.

Long, J. and J.S. Freese, 2006. *Regression Models for Categorical Dependant Variables using Stata, 2e.* College Station, TX. Stata Press.

Mellor, J.W., 1998. Agriculture on the Road to Industrialization, in Staatz and Eicher (eds) International Agricultural Development, 3e, Johns Hopkins University Press, Baltimore and London

Owango, M., B. Lukuyu, A., S.J. Staal, M. Kenyanjui, D. Njubi and W. Thorpe, 1998. Dairy Cooperatives and Policy Reform in Kenya: Effects of Livestock Service and Milk Market Liberalization. *Food Policy* Volume 23 (2), 173-185.

Staal, S. and G. Mullins, 1996. *Dairy Consumption and its Determinants in Coastal Kenya.* KARI/ILRI Collaborative Research Report. ILRI, Nairobi, Kenya.

Staal, S.J., C. Delgado and C. Nicholson, 1997. Smallholder Dairying Under Transaction Costs in East Africa. *World Development* Volume 25 (5), 779-794.

Thorpe, W., H.G. Muriuke, A. Omore, M.O. Owango and S. Staal, 2000. Dairy Development in Kenya: the past, the present and the future. Paper prepared for the Annual Symposium of the Animal Production Society of Kenya, Nairobi, Kenya. March 22 - 23, 2000.

Walshe, M. J., J. Grindle, A. Nell and M. Bachmann, 1991. Dairy Development in Sub-Saharan Africa: A Study of Issues and Options. World Bank Technical Paper, Number 135. Africa Technical Department Series. The World Bank, Washington DC.

Wooldridge, J.W. 2002. *Econometric Analysis of Cross Section and Panel Data.* Cambridge, Massachusetts: MIT Press.

World Bank. 2000. *Can Africa Claim the 21st Century?* Washington, DC.: World Bank.

World Bank. 2008. *Kenya Agricultural Policy Review: Current trends and future options for pro-poor agricultural growth.* Agriculture and Rural Development Unit, Discussion Paper.

World Economic Situation and Prospects. 2006. *World Economic Situation and Prospects 2006.* New York: Thu United Nations.

Yamano, T., and T.S. Jayne. 2004. *Working-age Adult Mortality and Primary School Attendance in Rural Kenya.* Working Paper No. 11. Nairobi, Kenya: Egerton University, Tegemeo Institute.