Liquid Milk

Geng, Xin; Janssens, Wendy; Kramer, B.N.

2017

INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE

IFPRI

IFPRI Discussion Paper 01602

January 2017

Liquid Milk

Cash Constraints and Day-to-Day Intertemporal Choice in Financial Diaries

Xin Geng

Wendy Janssens

Berber Kramer

Markets, Trade and Institutions Division

INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE

The International Food Policy Research Institute (IFPRI), established in 1975, provides evidence-based policy solutions to sustainably end hunger and malnutrition and reduce poverty. The institute conducts research, communicates results, optimizes partnerships, and builds capacity to ensure sustainable food production, promote healthy food systems, improve markets and trade, transform agriculture, build resilience, and strengthen institutions and governance. Gender is considered in all of the institute's work. IFPRI collaborates with partners around the world, including development implementers, public institutions, the private sector, and farmers' organizations, to ensure that local, national, regional, and global food policies are based on evidence.

AUTHORS

Xin Geng (x.geng@cgiar.org) is an associate research fellow in the Markets, Trade and Institutions Division of International Food Policy Research Institute (IFPRI), Washington, DC.

Wendy Janssens is an associate professor in the Department of Development Economics at VU University Amsterdam.

Berber Kramer (B.Kramer@cgiar.org) is a research fellow in the Markets, Trade and Institutions Division of IFPRI, Washington, DC.
Corresponding author

Notices

[1.] IFPRI Discussion Papers contain preliminary material and research results and are circulated in order to stimulate discussion and critical comment. They have not been subject to a formal external review via IFPRI's Publications Review Committee. Any opinions stated herein are those of the author(s) and are not necessarily representative of or endorsed by the International Food Policy Research Institute.

[2.] The boundaries and names shown and the designations used on the map(s) herein do not imply official endorsement or acceptance by the International Food Policy Research Institute (IFPRI) or its partners and contributors.

Contents

Tables

Figures

ABSTRACT

This paper analyzes implications of cash constraints for collective marketing, using the case of the Kenyan dairy sector. Collective marketing, for instance through cooperatives, can improve smallholder farmer income but relies on informal, nonenforceable agreements to sell outputs collectively. Side selling of output in the local market occurs frequently and is typically attributed to price differences between the market and cooperative. This paper provides an alternative explanation, namely that farmers sell in the local market when they are cash constrained because cooperatives defer payments while buyers in local markets pay cash immediately. Building on semiparametric estimation techniques for panel data, we find robust evidence of this theory. High-frequency, high-detail panel data show that farmers sell more in the local market, in particular to buyers who pay cash immediately, in weeks when they have low cash on hand. Moreover, households cope with health shocks by selling more milk in the local market and less to the cooperative, but only in weeks when they are not covered by health insurance. Increased flexibility in payment and the provision of insurance through agricultural cooperatives can potentially reduce side selling and improve the performance of collective marketing arrangements.

Keywords: liquidity constraints; side selling; panel data; semiparametric estimation

JEL codes: C23; D91; O16; Q13

ACKNOWLEDGMENTS

We are grateful to the PharmAccess Foundation and Amsterdam Institute for International Development (AIID) for funding the Health and Financial Diaries project and facilitating access to the data. We would like to thank Sicco van Gelder and Annegien Wilms (PharmAccess) for their valuable comments, Marijn van der List, David Pap and Lukasz Marc (AIID) for project and data management, and Hilda Kiritu and Hulda Kibera (Synovate Kenya) for their dedication to the collection of high-quality data. We highly appreciate the study participants' patience and time spent with enumerators throughout the study. Seminar participants at IFPRI provided useful suggestions and feedback. This work was undertaken as part of the CGIAR Research Program on Policies, Institutions, and Markets (PIM), which is led by IFPRI and funded by CGIAR Fund Donors. This paper has not gone through IFPRI's standard peer-review procedure. The opinions expressed here belong to the authors and do not necessarily reflect those of PIM, IFPRI, or CGIAR.

1. INTRODUCTION

In South Asia and Africa south of the Sahara, where the majority of farmers have small landholdings, low productivity in traditional agriculture is a major barrier to rural income growth. Although high-value agricultural commodities such as fruits, vegetables, meat, and milk provide an opportunity to increase income for smallholder farmers, their ability to market these commodities is challenged by high transaction costs, limited information regarding production methods and market opportunities, a lack of credit, and financial capital, low bargaining power and price volatility (Gulati et al. 2007). Collective marketing, for instance through cooperatives, can help relax some of these constraints, with significant welfare gains for smallholder farmers (Reardon et al. 2009; Minot and Sawyer 2014). Cooperatives can leverage economies of scale to reduce transaction costs, improve bargaining power vis-à-vis larger processing companies, and provide access to high-quality inputs as well as financial services.

To leverage economies of scale, cooperatives rely on their members to deliver reliable volumes of output. Many cooperatives therefore have bylaws or other informal agreements in place that require farmers to deliver their output to the cooperative. However, these agreements are typically nonenforceable. As a result, farmers often sell their output in local markets instead of delivering it to the cooperative, harming the performance of cooperatives. Literature on contract farming attributes such side selling to differences between cooperative and local market prices (Minot and Sawyer 2014). When cooperative prices drop below local market prices, for instance because of high local demand for the outputs, farmers are commonly assumed to deviate from the cooperative agreement and sell their output outside the cooperative.

This paper provides and tests an alternative explanation, namely that farmers' decisions on where to market their agricultural output are influenced by access to financial services, including access to savings accounts, insurance, and credit. Unlike traders and other buyers in local spot markets, cooperatives typically defer payments, which helps savings-constrained farmers accumulate savings; and farmers in fact do sell to cooperatives even when cooperative prices are lower than prices offered in local markets (Casaburi and Macchiavello 2015). Further, in the absence of formal insurance, smallholder farmers are exposed to uninsured risk, and they borrow at high interest rates to cope with unexpected cash needs (Collins et al. 2009). This is consistent with the theory that they face not only savings constraints but also credit constraints. We argue that in the absence of sound financial instruments to cope with shocks, farmers may sell their output in local markets when they are in financial need, even at lower prices than those offered by the cooperative.

We test this hypothesis using Health and Financial Diaries data, that is, high-frequency, high-detail panel data on household members' health and financial transactions, including milk production and consumption. We measure the degree to which a household is in need of cash by its lagged net nondairy income and lagged milk production. The analyses relate these variables to the share of milk delivered to the cooperative versus to other buyers (who pay immediately in cash), the share of dairy income from the cooperative versus to other buyers, and total dairy income. We also analyze how these predetermined variables affect the share of milk delivered to the cooperative versus other buyers, total dairy income, and the proportion of dairy income received on credit (as opposed to cash earnings from dairy farming).

Controlling for household and month-village fixed effects, both semiparametric and parametric regressions indicate that households receive a higher share of their dairy income from the cooperative in weeks that follow periods of higher levels of milk production and net income. In other words, a household appears to sell more milk to the cooperative and less to other buyers in weeks with more cash on hand. This effect is strongest after weeks of low milk production. We also observe an increase in side selling in the aftermath of a health shock, but only among households without health insurance coverage. These findings suggest that households use side selling as a coping strategy when they are in need of cash. However, we do not find evidence that this coping strategy lowers dairy income; although on average the cooperative offers a higher price than other buyers, farmers appear to sell more milk at a higher price when they are in need of cash.

This paper contributes to the existing literature in three ways. First, it models collective marketing decisions in the presence of both savings and credit constraints. Our model predicts that when cash on hand is high, savings-constrained farmers prefer selling to the cooperative to defer payments in order to save for future expenditures and to smooth consumption. However, when cash on hand is low, limited access to credit induces them to sell in local markets in order to obtain immediate cash. Independently, Casaburi and Macchiavello (2015) derived the first result. Nevertheless, their model did not analyze the effects of cash on hand and hence did not explain why the same farmer will either save with the cooperative or sell in local markets, depending on how cash constrained the farmer is at a given point in time. The present study therefore incorporates theory on how to interpret experimental measures of time preferences in a context of imperfect financial markets (Dean and Sautmann 2016; Epper 2015), showing the potential importance of corner solutions, threshold effects, and nonlinearities.

Second, this study contributes empirically by testing its theoretical predictions using high-detail, high-frequency panel data. It thereby analyzes real-life choices and demand for liquidity outside the stylized context of a laboratory. Casaburi and Macchiavello (2015) and Kramer and Kunst (2016) elicited demand for cash milk payments from cooperatives in the context of an experiment. This method provided higher control over observed choices but also came with the disadvantage that not all farmers would be cash constrained at the time of the experiment, and that the decision to sell to the experimenter—or the cooperative collaborating with the experimenter—is not the same as side selling to receive cash in the local market. By using yearlong panel data, the present study is able to observe households' behavior in weeks with more versus less cash on hand and test whether differences in cash on hand are related to subsequent milk marketing decisions. The finding that cash constraints matter in the decision of where to sell milk—and hence when to be paid—is consistent with recent field experiments showing that individuals with less wealth reveal less patient preferences (Dean and Sautmann 2016; Janssens, Kramer, and Swart 2016), but we show this for predictable, earned income allocations as opposed to unexpected windfalls or gifts.

A third contribution is that we test these predictions using semiparametric techniques that have to the best of our knowledge not been applied previously on microeconomic panel data. Semiparametric models include a nonparametric part with key explanatory variables as well as a parametric part with controls. Because such models do not presume any functional form for key explanatory variables, no potential misspecification biases the results. This characteristic is important given our theoretical prediction that net income affects the decision of where to sell milk differently depending on a household's level of milk production, and vice versa. In the semiparametric setting, both variables enter nonparametrically, which allows for a specification-free, pointwise, data-driven estimation of the individual and joint effects of these two variables. In addition, controlling for fixed effects in the parametric part minimizes an omitted variable bias. Compared with fully nonparametric regressions, this approach increases estimation precision and mitigates the "curse of dimensionality." We find that the semiparametric estimates often provide a more nuanced picture than a fully parametric model with a linear interaction term for lagged milk production and lagged net income, illustrating the use of semiparametric models when two variables have potential interaction effects.

The remainder of this paper is structured as follows. The next section presents a conceptual framework that will be used to derive the main hypotheses. Section 3. describes the data and semiparametric estimation method that will be used to test these hypotheses. Section 4. presents semiparametric results and compares those with those of a fully parametric linear estimator, followed by a number of robustness checks to further interpret the results. The final section provides a number of concluding remarks.

2. CONCEPTUAL FRAMEWORK

This section models collective marketing decisions in the presence of both savings and credit constraints. We model a setting in which farmers decide how much milk to deliver to the cooperative, which defers payment, and how much milk to sell in the local market, where milk is paid for immediately. We predict that when cash on hand is high, savings-constrained farmers prefer selling to the cooperative to defer payments as a savings device. However, when cash on hand is low, limited access to credit induces them to sell in local markets in order to obtain immediate cash.

Consider the following infinite-horizon framework with periods $t \in \{0, ..., \infty\}$. For every period, the farmer decides how much milk to sell in the local spot market, s_t, and how much to deliver to the cooperative, $m_t - s_t$, where m_t represents period-t milk production.[1] The local spot market provides dairy income immediately, in period t, while the cooperative does not pay until period $t+1$. Hence, period-t cash on hand is predetermined by nondairy cash income, y_t, as well as cooperative payments for milk delivered in the past, which depend on past milk production, m_{t-1}. We will analyze the effects of these two predetermined variables.

In period t, farmers face the following budget constraint:

$$c_t = y_t + p_t s_t + m_{t-1} - s_{t-1},$$

where c_t represents food consumption or expenditures in period t; y_t is nondairy cash income net of nonfood expenditures; s_t is the quantity of milk sold in the local market at a (relative) price of p_t; the cooperative milk price is normalized to 1, and $m_{t-1} - s_{t-1}$ is the quantity of milk delivered to the cooperative in the previous period, $t-1$, which is received as income in period t.

We assume that nondairy net income is predetermined. As a result, households can smooth food consumption, c_t, only by changing the quantity of milk sold to the cooperative versus to other buyers. Treating food consumption instead of nonfood expenditures as an endogenous variable is motivated by two factors. First, food consumption can vary substantially among cash constrained households, since these households have precautionary savings motives to smooth assets instead of consumption (Zimmerman and Carter 2003). Second, although empirically, cash constrained households may be able to increase their labor supply or forgo nonfood expenditures, the conceptual framework abstracts from this coping strategy. In the setting of interest, dairy farming and crop farming are the main sources of income, with limited opportunities for casual labor or other income-generating activities that provide cash on an emergency basis. Further, unavoidable spending on healthcare, school fees, and agricultural inputs such as seeds or animal feed constitutes a large share of nonfood expenditures. We hence model food consumption as an endogenous choice variable while including net income as a predetermined variable.

Another assumption implicit in the budget constraint is that the farmer has no access to savings devices outside the cooperative milk payments, and that borrowing is not possible. In other words, the farmer cannot save or borrow to smooth consumption; she can smooth consumption only by varying where she sells her milk. We impose this assumption for ease of representation, but it also reflects the notion that smallholder farmers have poor access to formal savings and credit instruments. Further, although households can receive transfers from their social network when in financial need, Ide et al. (2016) showed that our target population is unable to fully smooth consumption in the presence of health shocks, suggesting that informal insurance mechanisms are incomplete.

Given this budget constraint, farmers face the following decision-making problem:

$$\max_{0 \le s_t \le m_t} \sum_{t=0}^{\infty} \beta^t u\left(y_t + p_t s_t + m_{t-1} - s_{t-1}\right), \tag{1}$$

where $u(\cdot)$ is a standard increasing, concave, and twice-differentiable utility function, and β is an

[1] Note that by writing the quantity of milk delivered to the cooperative as $m_t - s_t$, we abstract from home consumption of milk and milk spillage.

exponential discount factor. Note that households utilize the quantity of milk delivered to the cooperative, $m_{t-1} - s_{t-1}$, as their only saving device. The theoretical predictions will focus on the effect of the two main components that drive a household's cash on hand - changes in current nondairy net income, y_t, and predetermined milk production, m_{t-1} - on the quantity of milk sold in the local market, s .

Table 2.1 summarizes the theoretical predictions for three equilibrium cases that exist under conditions specified in Lemma 2. (see Appendix A for a proof):

Lemma 1. Consider the utility maximization problem given in (1). An equilibrium exists in which (1) if $\bar{p} < \beta$, the farmer sells all milk to the cooperative, that is, $\bar{s} = 0$; (2) if $\bar{p} > \beta$, the farmer sells all milk in the local spot market, that is, $\bar{s} = \bar{m}$; (3) if $\bar{p} = \beta$, the farmer is indifferent about where to sell, that is, $\bar{s} \in (0, \bar{m})$.

Table 2.1 Theoretical predictions on how cash on hand affects buyer choice

Case 1: $\bar{s} = 0$ $(\bar{p} < \beta)$			
a) Increase in current nondairy income, past milk production, or both			
$y_t > \bar{y}$,	$m_{t-1} > \bar{m}$	$\implies$	$s_t = 0$
b) Decrease in current nondairy income, past milk production, or both			
$y_t < \bar{y}$,	$m_{t-1} < \bar{m}$	$\implies$	$s_t > 0$
Case 2: $\bar{s} = \bar{m}$ $(\bar{p} > \beta)$			
a) Increase in current nondairy income, past milk production, or both			
$y_t > \bar{y}$,	$m_{t-1} > \bar{m}$	$\implies$	$s_t < \bar{m}$
b) Decrease in current nondairy income, past milk production, or both			
$y_t < \bar{y}$,	$m_{t-1} < \bar{m}$	$\implies$	$s_t = \bar{m}$
Case 3: $\bar{s} \in (0, \bar{m})$ $(\bar{p} = \beta)$			
a) Increase in current nondairy income, past milk production, or both			
$y_t > \bar{y}$,	$m_{t-1} > \bar{m}$	$\implies$	$s_t < \bar{s}$
b) Decrease in current nondairy income, past milk production, or both			
$y_t < \bar{y}$,	$m_{t-1} < \bar{m}$	$\implies$	$s_t > \bar{s}$

Source: Summary of results from Lemma 1–4.
Note: $\bar{x}$ stands for the equilibrium level of variable x, $x = y_t, m_t, s_t$.

In the first case, the local market price is below the discounted cooperative price. As a result, farmers sell all milk to the cooperative in equilibrium. In the second equilibrium case, the local market offers a lower price than the cooperative does, but because it pays farmers sooner, at a price that is above the discounted cooperative price, farmers prefer to sell all milk in the local market. In the third case, the price offered by the local market equals the discounted price that the cooperative pays. As a result, farmers are indifferent in equilibrium as to whether they sell the milk in the local market or to the cooperative.

Appendix A derives the response to changes in current income, y_t, and past milk production, m_{t-1}, for each of these three cases. In the first case, the cooperative offers a relatively high price, even when taking into account farmers' discount rates. As a result, farmers prefer selling all milk to the cooperative instead of selling it in the local market. The quantity of milk sold in the local market hence does not change with an increase in (y_t, m_{t-1}); both result in an increase in cash on hand, reducing farmers' demand for sooner payments, which they could obtain from the local market. In contrast, a decrease in (y_t, m_{t-1}) *can* induce farmers to sell in the local market because they have less cash on hand than expected and need more cash in order to smooth consumption. Appendix A shows that reductions in current income and past milk production have this effect under the following condition:

Lemma 2. Consider the utility maximization problem given in (1) in the equilibrium when $\bar{p} < \beta$ such that $\bar{s} = 0$. Then, (1) if there is an increase in current income, $y_t > \bar{y}$; an increase in past milk production, $m_{t-1} > \bar{m}$; or both, then $s_t = 0$; (2) if there is a decrease in y_t, m_{t-1}, or both, such that $u'(y_t + m_{t-1})/u'(\bar{y} + \bar{m}) > \beta/\bar{p}$, then $s_t > 0$. Further, if the response s_t satisfies $u'(\bar{y} + \bar{m} - s_t)/u'(\bar{y} + \bar{m}) > \beta/\bar{p}$, then also $s_{t+1} > 0$.

Intuitively, if cash on hand, $y_t + m_{t-1}$, is sufficiently low relative to the equilibrium level, $\bar{y} + \bar{m}$, so that the marginal rate of substitution between the current and next periods exceeds the threshold $\beta/\bar{p}$, then farmers will sell milk in the local market. This threshold is increasing in the local market price, meaning that as this price drops further below the cooperative price, cash on hand needs to drop by a more substantial amount in order to induce farmers to sell in the local market. The effect of a drop in y_t or m_{t-1} can perpetuate into the future because in the next period, milk payments from the cooperative—and hence cash on hand—will again be below the equilibrium level. As a result, farmers will be induced to sell a strictly positive, but decreasing, quantity of milk in the local market for a prolonged period of time, limiting the quantity of milk sold to the cooperative.

We find a similar result, although in the opposite direction, in Case 2. When farmers sell all milk in the local market, a sufficient increase in (y_t, m_{t-1})—and hence an increase in cash on hand—reduces demand for immediate cash, making it optimal to sell some milk to the cooperative. Appendix A shows that this occurs under the following condition:

Lemma 3. Consider the utility maximization problem given in (1) in the equilibrium when $\bar{p} > \beta$ such that $\bar{s} = \bar{m}$. Then, (1) if there is a decrease in current income, $y_t < \bar{y}$; a decrease in past milk production, $m_{t-1} < \bar{m}$; or both, then $s_t = \bar{m}$; (2) if there is an increase in y_t, m_{t-1}, or both, such that $u'(y_t + \bar{p}\bar{m} + m_{t-1} - s_{t-1})/u'(\bar{y} + \bar{p}\bar{m}) < \beta/\bar{p}$, then $s_t < \bar{m}$. Further, if the response s_t satisfies $u'(\bar{y} + \bar{p}\bar{m} + \bar{m} - s_t)/u'(\bar{y} + \bar{m}) < \beta/\bar{p}$, then also $s_{t+1} < \bar{m}$.

In this case, the cooperative pays a higher price than the local market, but the higher price cannot compensate the farmer sufficiently for the strong discounting of deferred payments. Hence, in equilibrium, the farmer prefers selling in the local market. Nonetheless, an increase in current income or milk production increases the amount of cash on hand compared with future periods. Instead of spending all cash on consumption in the current period, farmers will prefer to save and smooth consumption over time by selling some milk to the cooperative. This is why not only current income but also an increase in lagged milk production results in higher cash on hand in the current period, and as in the previous case, shocks to income or milk production can hence have a prolonged effect.

In the third and final case, farmers are indifferent as to where to sell the milk. Now, the quantity sold in the local market (s_t) will respond inversely to the change in (y_t, m_{t-1}): an increase in current income or milk production increases the amount of cash on hand and hence lowers the demand for sooner payments, resulting in increased quantities of milk sold to the cooperative, while a decrease in income or milk production lowers the amount of cash on hand, resulting in a higher demand for sooner payments. This also affects the amount of cash on hand in future periods, resulting in perpetuating effects, but of a smaller magnitude.

Lemma 4. Consider the utility maximization problem given in (1) in the equilibrium when $\bar{p} = \beta$ such that $\bar{s} \in (0, \bar{m})$. Then, (1) if there is a decrease in current income, $y_t < \bar{y}$; a decrease in past milk production, $m_{t-1} < \bar{m}$; or both, then $\bar{s} < s_t < s_{t+1} < \cdots$, but $\{s_t\}$ does not necessarily converge to $\bar{m}$, depending on the property of the underlying utility function; (2) if there is an increase in y_t, m_{t-1}, or both, then $\bar{s} > s_t > s_{t+1} > \cdots$, and $\{s_t\}$ does not necessarily converge to 0.

The above analysis also indicates that in theory, our outcome variables are related to lagged milk production and current income in a nonlinear way. As long as the relative milk price in the local market does not equal the discount factor, changes in net income or lagged milk production have to pass a certain threshold in order to affect the quantity of milk sold in the local market.[2] This suggests that s_t is likely to be a piecewise function of the two indicators for cash on hand. In addition, Appendix A shows that this relation is piecewise linear only when we impose a quadratic, nonconcave utility function, which is an unlikely empirical scenario.

Although we focus on the case in which local markets offer lower prices than the cooperative offers, $\bar{p} < 1$, this argument generalizes to a situation in which the cooperative pays a lower price than the local market, $\bar{p} > 1$. The quantity of milk sold in the local market will generally respond in a similar way, and the conditions derived in Appendix A will not change.

[2] Appendix A gives thresholds for each case.

3. METHODS

Context and Data

We test our theoretical predictions empirically using high-detail, high-frequency panel data collected over a year for a sample of 120 dairy farmers and their families. All farmers were members of a dairy cooperative in Nandi County. This county, near Eldoret in western Kenya, has a population of 752,965 individuals, and agriculture is the main economic activity. Dairy farming is an important income source for a large proportion of the population. Other than this, Nandi County is an ideal area for tea, maize, and sugarcane farming. The poverty rate in the county is 47.4 percent, which is about equal to the country's average. The share of the population living in urban areas is 13.6 percent and approximately 67.3 percent of the population has completed primary education.

The dairy cooperative, named Tanykina, collects milk through several milk transporters and collection centers.[3] Other than collectively marketing milk, Tanykina sells agricultural inputs, such as animal feed and fertilizers, as well as veterinary services. Farmers can pay for these inputs through their milk account. Further, through its Savings and Credit Cooperative Organization (SACCO), Tanykina provides financial services including agricultural loans and savings accounts. In practice, these services are likely to be underutilized. For a similar sample of dairy farmers, Jack et al. (2016) found a very low 2.4 percent take-up of SACCO loans with comparable stringent borrower requirements. Further, these loans are mainly used for agricultural investments, not to cope with financial shocks such as health expenditures or to make nonagricultural expenditures. Savings are mainly used to meet deposit requirements for accessing loans.

Finally, at the time of the study, Tanykina members could enroll in a health insurance plan named The Community Health Plan (TCHP). At the time of the study, TCHP was offered by AAR Insurance, a Kenyan insurance provider, with support from the PharmAccess Foundation and Health Insurance Fund.[4] Ide et al. (2016) described TCHP in more detail and showed that it provides financial protection from health shocks and results in improved food security, in particular among households with weaker social networks.

Larger milk processing companies collect milk from Tanykina collection centers on a daily basis. However, to reduce transaction costs, they pay Tanykina only once per month, after assessing the actual quantity of milk collected from the cooperative in a given month. Once the processor has determined how much Tanykina will receive, the cooperative can start preparing farmers' milk payments. Due to this process, Tanykina defers milk payments until the month following delivery. In preparing farmers' milk payments, Tanykina takes into account the total quantity of milk a farmer delivered in a given month as well as any deductions for shares; inputs purchased from Tanykina; and the costs of any services utilized, including TCHP insurance premium.

The study selected three dairy collection areas to implement data collection: Salien, Surungai, and Lemook. From these collection areas, the study randomly selected seven villages with a minimum of 25 Tanykina member households: three from Salien, two from Surungai, and two from Lemook. The study randomly selected 120 Tanykina member households from these seven villages. The number of selected Tanykina members within a village is proportional to the total number of Tanykina members in that village. Further, the sample is stratified by insurance status, with half of the households having insurance at baseline. The proportion of insured versus uninsured households sampled within a village is proportional to the total number of insured households in the village.

The data were collected between October 2012 and October 2013 as part of the Health and Financial Diaries project (Janssens et al. 2013). The goal of this yearlong study was to gain

[3] Formally, Tanykina Dairies Ltd. is not a cooperative but a private company limited by shares. Because its shares are owned by farmers, and because otherwise the company acts as a cooperative, we will henceforth refer to the company as Tanykina or "the cooperative."

[4] Initially the program targeted Tanykina members and their families, but it has transitioned to being accessible to the general public.

understanding of health-seeking behavior and financial lives in the TCHP target population. For the 120 selected households, enumerators interviewed all 207 financially active household members, both males and females, separately and in private. They did so every week for the duration of a year. Respondents were interviewed about all financial transactions since the last interview including their savings, gifts, loans, income, and expenditures. For each transaction, the diaries collected information on the transaction item ("what"), the date ("when"), the transaction partner ("who"), the transaction mode ("how"; for example, whether a payment was in cash, via mobile money, or on credit), and the transaction value ('how much'). We use these records to identify financial transactions from selling milk to the cooperative versus other buyers in the local market.[5] In addition, the diaries collected information about family members' health, milk production, and milk consumption. Because these data are recorded at the household level, we treat households instead of individuals as the unit of observation. We will therefore aggregate individuals' financial transactions at the household level as well.

Despite the high intensity of data collection, attrition was low; only two households dropped out during the study. The goal was to interview the remaining households every week throughout the year. This plan was feasible in the vast majority of weeks, with the exception of the two weeks around Christmas; the week of the elections, when no data collection occurred for security reasons; and the final month of data collection. We exclude these weeks from the analyses. We also exclude household i in week t if no household member was available for an interview in week t. This was the case for 7.3 percent of all potential household interview weeks. If not all but at least one household member completed an interview in week t, and if this person reported selling milk, we do include this observation. Because data on milk sales and health, as well as milk production and consumption (collected at the household level), are available for these weeks, only the individual financial transactions for absent household members are missing for these weeks. To avoid omitting these weeks, we impute income and expenditures for an absent respondent by his or her yearly average.

The analyses will also use a baseline survey, which was completed with all respondents prior to the start of the Health and Financial Diaries, and a separate dataset containing information on monthly enrollment, renewal, and suspension of the individuals in our sample in the TCHP program. For members deciding to enroll their families in TCHP, the premium was deducted automatically from the monthly milk payment. In case of insufficient funds, the farmer could pay the premium in cash or through mobile money. If not paying the monthly premium, a household was suspended the next month and could not benefit from the insurance policy; if the household paid a double premium in the next month, its insurance coverage was reinstated; otherwise, the household was dropped from TCHP for the rest of the year. We consider a household to have insurance coverage if and only if at least one family member has (nonsuspended) insurance coverage.

Measurement of Main Variables Used in the Analysis

We empirically analyze farmers' milk marketing behavior in terms of three outcome variables: the share of milk sold to the cooperative ($ShareMilkTan \equiv 1 - s_t/m_t$), total dairy income ($SalesAll \equiv m_t - s_t + p_t s_t$),[6], and the share of dairy income from the cooperative ($ShareTan \equiv 1 - p_t s_t/SalesAll$), where m_t, s_t, and p_t, defined in Section 2. represent current milk production net of milk consumption, the quantity of milk sold outside the cooperative, and the market milk price relative to the cooperative price, respectively. Cooperatives are interested in maximizing the share of milk sold to the cooperative, *ShareMilkTan*, while farmers will be more interested in maximizing their total dairy income, *SalesAll*. Policy makers aiming at strengthening the formal dairy value chain through cooperatives will want to maximize *ShareTan* because it takes into account not only the quantity of milk delivered to the cooperative but also the price offered by the cooperative relative to the local market price.

[5] The questionnaires recorded the value of every transaction but not the number of items or the price per item. As a result, we do not have primary data on milk prices and the quantity of milk sold at the household level.

[6] We apply a log transformation to reduce the influence of outlier values.

Every week, we observe the fixed monthly cooperative price (used as numeraire in our conceptual framework); the quantity of milk production net of milk consumption, m_t; and the total value of milk sold to the cooperative, $m_t - s_t$ (again using cooperative prices as the numeraire), and to other buyers, $s_t p_t$.[7] These earnings are used to directly measure the share of dairy income from the cooperative, *ShareTan*, and total dairy income, *SalesAll*. However, we do not directly measure the *quantity* of milk sold to the cooperative as a share of the total quantity of milk sold, *ShareMilkTan*. Households did not report how much milk they sold to a given buyer, or at what price. Although data collection included a monthly market survey on milk prices, local market prices may vary over time and by household. For this reason, we estimate the share of milk sold to the cooperative as follows. First, to derive the quantity of milk sold to the cooperative, we divide earnings from the cooperative by its monthly milk price. We then divide this quantity by total milk production net of milk consumption (that is, the total quantity of milk sold, m_t) in order to obtain the share of milk that is sold to the cooperative.

The share of milk sold to the cooperative, *ShareMilkTan*, and the share of dairy income from the cooperative, *ShareTan*, are both decreasing in the milk quantity sold in the local market, s_t. Further, under the condition that the cooperative offers a higher price than buyers in the local market, $p_t < 1$, total dairy income, *SalesAll*, is also decreasing in s_t. In our empirical application, the cooperative offers higher prices than the average buyer in local markets. In that case, the choice of whether to sell milk to the cooperative or in the local market involves a trade-off between later-larger and smaller-sooner rewards. Given that in this scenario, the quantity of milk sold in the local market, s_t, is negatively correlated with each of our outcome variables, the theoretical predictions from the conceptual framework indicate that all of our outcome variables are increasing with current nondairy net income, y_t, and predetermined milk production, m_{t-1}, with potential threshold effects and nonlinearities. The only caveat to keep in mind is that when the cooperative offers a lower price than local markets, selling milk to the cooperative will not optimize total dairy income. In that case, the lack of alternative savings instruments is potentially the reason that households prefer deferred payments from the cooperative even when deferred payments offer a lower price than immediate payments, as observed by Casaburi and Macchiavello (2015) and Kramer and Kunst (2016).

Finally, we proxy households' cash on hand by two key explanatory variables: log milk production in liters (*L2MilkProd*) and log nondairy cash income minus log nonfood expenditures (*L2NetInc*), both in 1,000 Kenyan shillings (KSh).[8] Both variables are averaged over the past two weeks and hence predetermined. In this way, our results will not be driven by reverse causality.

Fully Linear Model

To study the individual and joint effects of *L2MilkProd* and *L2NetInc* on each of the main outcome variables, we can estimate a simple fully linear model with an interaction term and fixed effect:

$$y_{it} = \alpha_i + L2MilkProd\gamma_1 + L2NetInc\gamma_2 + L2MilkProd \times L2NetInc\gamma_3 + x_{it}'\beta + v_{it}, \\ \text{i} = 1, \cdots, n, \quad t = 1, \cdots, T, \tag{2}$$

where y_{it} is one of the three outcome variables, i indexes household, t indexes week, and α_i is a household fixed effect. This estimating equation is linear in the predetermined variables *L2MilkProd*, *L2NetInc*, and their interaction term, as well as x_{it}, a $p \times 1$ vector of control variables. Across specifications, we vary the variables included. x_{it} might include current milk production (*MilkProd*) and milk consumption (*MilkCons*), and always includes the following predetermined variables averaged over the past two weeks: log food expenses in KSh 1,000 (*L2Food*), whether anyone in the household incurred a health problem (*L2HealthProblem*), whether at least one household member is

[7] Although households report their expected dairy income from Tanykina weekly, it is paid on a monthly basis. The total value of milk sold to the cooperative is hence not a perfect indicator of cash inflows from the cooperative in the next period.

[8] The value of KSh 1,000 was approximately US $11.50 at the time of data collection.

insured (*L2Insured*), and the interaction of the last two variables.[9] β is a $p \times 1$ vector of unknown parameters, and v_{it} is an independent and identically distributed (i.i.d.) error term.

The specification described in Equation (2) is not based on theoretical economic foundations but is a good candidate for a reduced form to estimate how different sources of cash on hand affect farmers' preference on where to sell their milk. At the same time, the effects of cash on hand may be piecewise and nonlinear, and given the inclusion of an interaction term, estimates for γ_1 and γ_2, that is, the individual effects of *L2MilkProd* and *L2NetInc*, respectively, will depend crucially on the level at which the other variable is centered. If this model setup is not correct (most likely so), the estimates could be subject to great bias and inaccurate inference. This situation calls for a more flexible, nonparametric characterization.

Semiparametric Model and Estimation

We estimate a semiparametric model to investigate how predetermined indicators of cash on hand relate to farmers' marketing behavior. The nonparametric part of this model includes the main variables used to proxy cash on hand—lagged milk production (in logs) and lagged nondairy income net of nonfood expenditures (both in logs). The parametric part, which is linear, controls for other variables, including household fixed effects.

The contribution of this semiparametric model is that it allows for a specification-free, pointwise, data-driven estimation for the nonparametric part. This capability is important given the prediction of threshold effects and nonlinear response functions in Section 2. Moreover, in a fully linear model with an interaction term, the individual effects of the two predetermined variables are presumed to be linear in the other variable and are sensitive to the level at which the other variable is centered, while the semiparametric estimates are not subject to this constraint. Thus, a semiparametric model can yield a better description of how the variables included in the nonparametric part affect the quantity of milk sold in the local market. This also increases the precision of the estimated coefficients for variables included in the linear part.

Note that the first case in the previous section, $\bar{p} < \beta$, modeled lagged milk production and net income as perfect substitutes, with current income resembling cash inflows from nondairy activities, and lagged milk production resembling the cash payment from the cooperative for milk delivered in the previous period. Empirically, we cannot treat these variables as perfect substitutes because they are not necessarily perfectly fungible. Farmers are unlikely to be in the first-case equilibrium because most households sell at least some milk in the local market on a regular basis. Hence, lagged milk production enters the current budget constraint as dairy income only after deducting any milk sold in the local market in the previous period. Unlike current nondairy income, past milk production is not a direct measure of cash flowing in from dairy activities, even if we were to scale this quantity by the cooperative price that the previous section uses as numeraire. Further, we cannot include milk sold in the local market in the previous period to get a better measure of predetermined cash flows from dairy farming, given that this is a lagged dependent variable. In addition, dairy income from the cooperative received in cash is subject to larger measurement error than nondairy income because it is delivered only once a month and we do not measure its exact timing. Finally, for changes in lagged milk production, there is a potential compensating effect from adjusting the quantity of milk sold in the local market in the previous period. For current net cash inflows from nondairy activities, such compensating effects do not occur.

The foregoing implies that the empirical strategy need not only to account for nonlinearities and threshold effects of cash on hand but also to be applicable to a setting in which such nonlinearities and threshold effects occur for two variables that potentially interact with one another, given that they are not perfect substitutes. We therefore consider the following partially linear model with fixed effects,

[9] We use an indicator for whether at least one member of the household has insurance coverage. Thus, there is a possibility that the sick individual is not insured, even though the household is recorded as insured.

which was originally proposed by Su and Ullah (2006):

$$y_{it} = \alpha_i + m(z_{it}) + x_{it}'\beta + v_{it}, \quad i = 1, \cdots, n, \quad t = 1, \cdots, T. \tag{3}$$

Here, y_{it}, the dependent variable, is again one of our three variables of interest: the share of milk sold to Tanykina (*ShareMilkTan*), the share of dairy income from Tanykina (*ShareTan*), and total dairy income in a given week (*SalesAll*). Further, z_{it}, a vector of two variables (*L2MilkProd* and *L2NetInc*), enters nonparametrically because $m(\cdot)$ is an unknown smooth function, rather than linearly with an interaction term in the linear setting. x_{it} includes the same control variables as in Equation (2), and v_{it} is the random disturbance. We always control for month-village fixed effects that will capture spatial variation over time in unobserved characteristics, including the difference between local market and cooperative milk prices. The estimation also always includes household fixed effects.

For identification, assume $\sum_{i=1}^{n} \alpha_i = 0$,[10] and $\mathrm{E}(v_{it}|x_i, z_i, \alpha_i) = \mathrm{E}(v_{it}|x_{it}, z_{it}) = 0$, where $x_i = (x_{i1}, \cdots, x_{iT})'$, and z_i is analogously defined.

Su and Ullah (2006) proposed consistent estimators for β, $m(\cdot)$, and $\dot{m}(\cdot)$ for an i.i.d. process, where $\dot{m}(\cdot)$ is the first derivative of $m(\cdot)$, and they established the asymptotic properties with fixed T and n going to infinity. The parameter estimation is based on the profile likelihood method, specifically, profile least squares. This method builds on the idea that if the linear parameters were given, we could rearrange Equation (1) and obtain an unfeasible but explicit estimator for the nonparametric part, with $y_{it} - \alpha_i - x_{it}\beta$ as the new dependent variable. The nonparametric estimator is a function only of the unknown linear parameters. Substituting $m(z_{it})$ for this nonparametric estimator in the original estimating equation, we can rearrange again such that we obtain the parametric estimators using traditional ordinary least squares. The feasible nonparametric estimator, given the parametric estimator, then follows immediately.

To conceptualize this estimation process, let $\alpha = (\alpha_2, \cdots, \alpha_n)'$, $D = (I_n \otimes i_T)d_n$, and $d_n = (-i_{n-1}, I_{n-1})'$. We can then write Equation (3) in matrix form:

$$Y = D\alpha + M + X\beta + v, \tag{4}$$

where $Y = (y_{11}, \cdots, y_{1T}, y_{21}, \cdots, y_{nT})'$, $X = (x_{11}, \cdots, x_{1T}, x_{21}, \cdots, x_{nT})'$, $v = (v_{11}, \cdots, v_{1T}, x_{21}, \cdots, v_{nT})'$, and $M = (m(z_{11}), \cdots, m(z_{1T}), m(z_{21}), \cdots, m(z_{nT}))'$.

Using this matrix form, the estimation process proceeds as follows:

1. If α and β were known, we would have a purely nonparametric regression:

$$Y - D\alpha - X\beta = M + v.$$

2. Obtain the local linear estimator for $M(z) \equiv \big(m(z), (H\dot{m}(z))'\big)'$ that includes $m(z)$:

$$M_{\alpha,\beta}(z) = S(z)(Y - D\alpha - X\beta), \qquad m_{\alpha,\beta}(z) = s(z)'(Y - D\alpha - X\beta), \tag{5}$$

where $S(z)$ and $s(z)$ are the local linear premultipliers, explicitly defined in Appendix B.

3. Apply the unfeasible estimator $m_{\alpha,\beta}(z_{it})$ for $m(z_{it})$ in (4) and rearrange. By partitioned regression formula, we obtain the least squares parametric estimators:

$$\hat{\beta} = (X^{*\prime}M^*X^*)^{-1}X^{*\prime}M^*Y^* \quad \text{and} \quad \hat{\alpha} \equiv (\hat{\alpha}_2, \cdots, \hat{\alpha}_n)' = (D^{*\prime}D^*)^{-1}D^{*\prime}(Y^* - X^*\hat{\beta}), \tag{6}$$

where $A^* \equiv (I_{nT} - S)A$, for $A = Y, X, D$; $S \equiv (s_{11}, \cdots, s_{1T}, s_{21}, \cdots, s_{nT})'$ and $s_{it} \equiv s(z_{it})$.

[10] We do not include a constant term in Equation (3) because we would be unable to identify a constant separately from the unknown function $m(\cdot)$ unless we constrained $m(\cdot)$ to have, for example, a 0 mean. Because the sum of the n household fixed effects is 0 (and hence cannot capture any of the sample mean), the nonparametric part, $m(\cdot)$, and the other linear part will capture the sample mean.

4. Plugging $\hat{\beta}$ and $\hat{\alpha}$ back into (5), we have the estimators for $M(z)$; and $m(z)$:

$$\hat{M}(z) = S(z)\big(Y - D\hat{\alpha} - X\hat{\beta}\big) \quad \text{and} \quad \hat{m}(z) = s(z)'\big(Y - D\hat{\alpha} - X\hat{\beta}\big). \tag{7}$$

The estimator for α_1 follows as $\hat{\alpha}_1 = -\sum_{i=2}^{n} \hat{\alpha}_i$.

In Appendix B, we provide a brief introduction to the local linear estimation and asymptotic properties of $\hat{\beta}$ and $\hat{M}(z)$, based on which we construct the confidence intervals. We also give details regarding the implementation of the nonparametric estimation, including the bandwidth selection procedure. Consistent estimators of the covariance matrices are also given. We refer readers to Su and Ullah (2006) for detailed proofs.

Sample Description

Table 3.1 presents an overview of household characteristics at the start of the financial diaries, omitting the two households that dropped out of the study. Because our model includes household fixed effects, identification relies on variation in our outcome variable within households over time. Our main outcome variable is the proportion of dairy income received from Tanykina, the cooperative, and the analyses are hence restricted to the 88 households with variation in this variable.

Table 3.1 Summary statistics of household characteristics

	Variation in share of income from Tanykina		No variation in share of income		Difference in means	
Variable	Mean (1)	s.e. (2)	Mean (3)	s.e. (4)	Mean (5)	*p*-value (6)
Household head is male	0.705	0.459	0.500	0.509	0.205	0.0425
Age of the household head	52.38	14.15	51.03	19.13	1.342	0.6839
Number of HH members selling milk	1.489	0.547	1.300	0.466	0.189	0.0935
Number of cows at baseline	4.227	2.509	3.200	1.669	1.027	0.0390
Number of cows lent out at baseline	0.125	0.333	0.069	0.258	0.056	0.4094
Number of bulls at baseline	0.356	0.849	0.333	0.606	0.023	0.8916
Number of male calves at baseline	1.227	1.444	0.833	1.234	0.394	0.1841
Number of female calves at baseline	3.318	10.22	1.233	1.073	2.085	0.2684
Number of weeks with an interview	45.42	4.085	46.03	2.141	-0.613	0.4344
Main dairy farmer:						
Is male	0.216	0.414	0.300	0.466	-0.084	0.3541
Age	47.57	14.34	46.40	17.38	1.168	0.7161
Is household head	0.477	0.502	0.733	0.450	-0.256	0.0148
Is spouse of household head	0.500	0.503	0.200	0.407	0.300	0.0038
Is married	0.830	0.378	0.600	0.498	0.230	0.0095
Went to secondary school or higher	0.531	0.502	0.423	0.504	0.108	0.3435
Is Protestant	0.830	0.378	0.933	0.254	-0.104	0.1649
Is engaged in livestock activities	0.953	0.213	0.967	0.183	-0.014	0.7539
Is engaged in cattle activities	0.765	0.427	0.833	0.379	-0.069	0.4378
Is engaged in other livestock activities	0.635	0.484	0.567	0.504	0.069	0.5104
Can keep part of cattle income	0.659	0.477	0.793	0.412	-0.134	0.1789
Decides how to spend cattle income	0.655	0.478	0.793	0.412	-0.138	0.1677
Exp. cattle income next month in 1,000 KSh	1.821	2.189	6.272	21.63	-4.451	0.0620
Exp. cattle income next year in 1,000 KSh	26.77	35.87	67.01	188.7	-40.23	0.0619
Total exp. income next month in 1,000 KSh	6.167	13.00	10.52	22.53	-4.356	0.2061
Total exp. income next year in 1,000 KSh	158.0	282.7	225.7	314.4	-67.72	0.2758
Number of households	88		30			

Source: Data from the Health and Financial Diaries project in Kenya (Janssens et al., 2013).
Note: The sample excludes two households with high attrition and strong outliers. Further, in assessing whether there is variation over time in the share of income received from Tanykina, we omit weeks with high attrition due to Christmas (two weeks), Easter (one week) and the last fieldwork month (four weeks). The main dairy farmer is the household member reporting the highest value of dairy income received throughout the year. HH = household; KSh = Kenyan shillings.

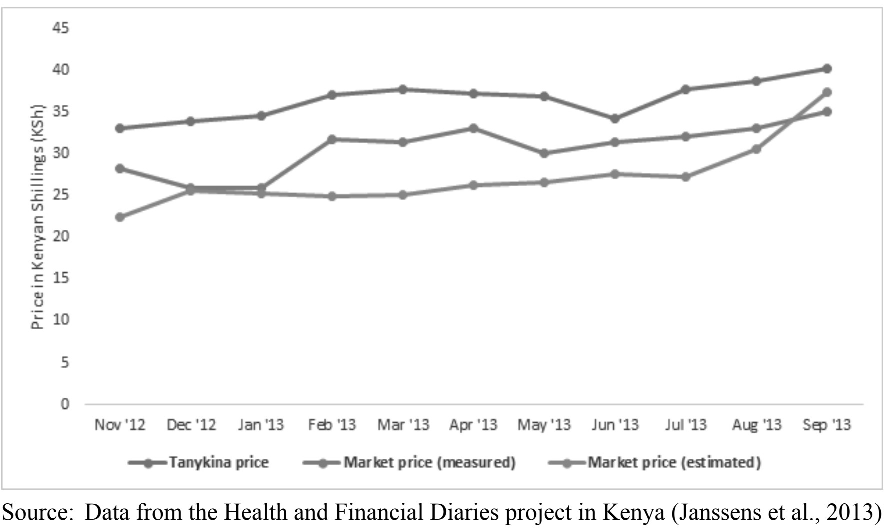
Columns (1)–(2) describe the analysis sample that satisfies this condition. In 70.5 percent of the households, the main decision maker is male. The average age of the household head is 52.4 years. Around 1.5 members of the average household reported selling milk at some point during the Health and Financial Diaries. These households have on average 4.2 cows. Households are much more likely to keep cows than they are to keep bulls. They have around 4.5 calves, of which the majority are female as well.

In half of the sample, the main household member selling milk is not the (male) household head but his wife. She is on average five years younger than the household head. Around half of all dairy farmers completed at most primary education and never went to secondary school. The sample is mostly Protestant. At baseline, 95.3 percent of main dairy farmers reported being engaged in livestock activities, but not all of them reported dairy farming at that time. If engaged in dairy farming, 86.1 percent keep some of the dairy income, and nearly all can decide how to spend that money. Expected dairy income is KSh 1,821 (approximately US $18.21) in the next month, or 29.5 percent of total expected income for the main dairy farmer. Over the year, expected dairy income is KSh 26,770 (approximately US $267.70), or 16.9 percent of total expected income.

Columns (3)–(4) describe households without variation in the share of income from Tanykina, and Columns (5)–(6) compare this sample with the sample of households in the analyses. Households without variation in the share of income from Tanykina either always sell to the cooperative or always sell in the local market. These households are more likely to have a female household head, have fewer household members selling milk, and have on average one cow less at baseline. Their main dairy farmer is significantly more likely to be the household head instead of his spouse, and this farmer is less likely to be married. In these households, dairy farming makes up a higher share of household income; expected income from dairy farming is higher, while this is not the case for total income.

Figure 3.1 plots the milk price offered by Tanykina and the milk price in the local market over time. The milk price in the local market was collected through a monthly survey of market prices. For each commodity, the interviewer would ask three market vendors at what price they currently sell that commodity. Figure 3.1 draws for every month the average price. It is likely that farmers receive a lower price for their milk than the price at which vendors sell their milk, making this price an upper bound for the actual average market price. Nonetheless, the price offered by the cooperative is higher than the average price in the local market throughout the study period. The difference ranges between KSh 3 (in June) and KSh 9.5 (in January), which is 8.3 and 26.2 percent of the average Tanykina price, respectively.

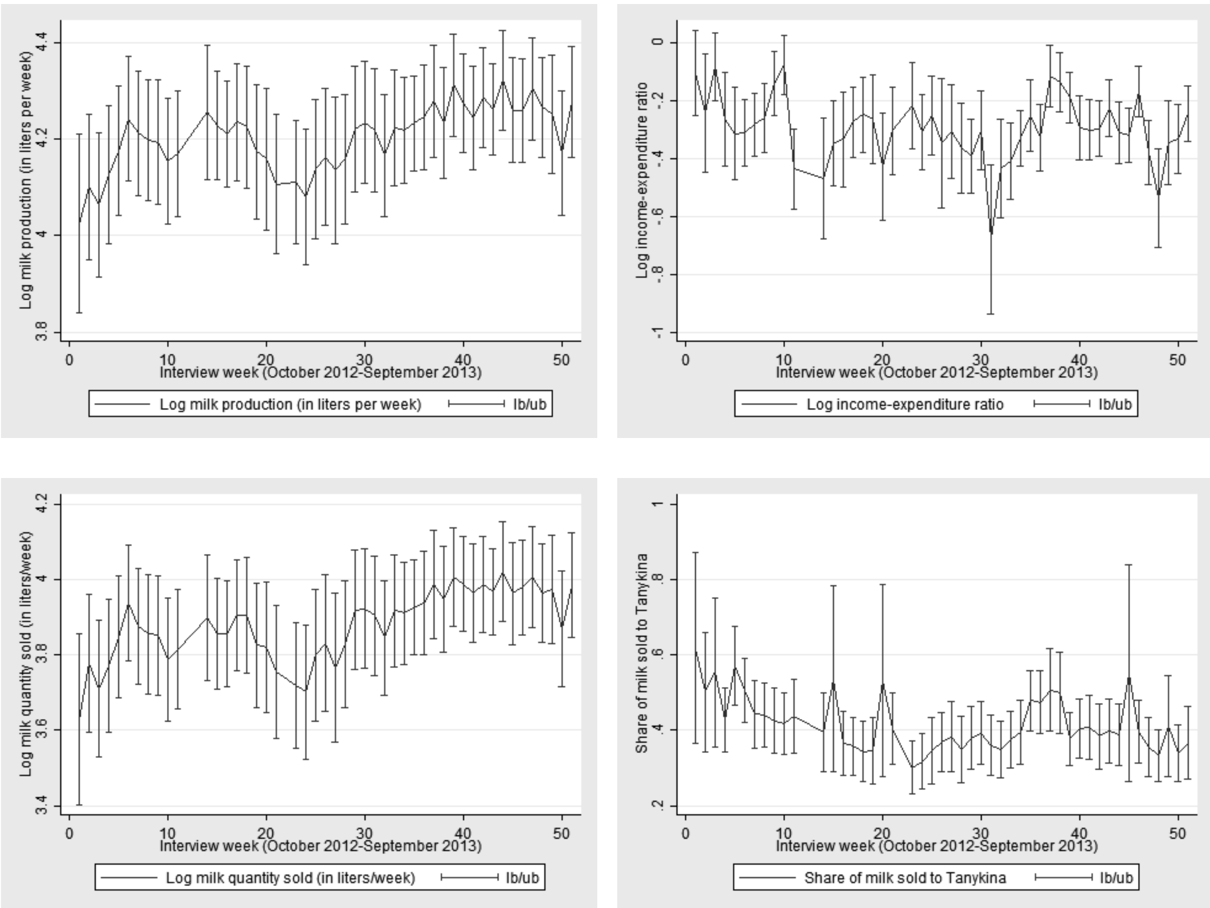
Figure 3.1 Milk price paid by Tanykina versus average milk price in local market



Source: Data from the Health and Financial Diaries project in Kenya (Janssens et al., 2013).
Note: Time series of fixed monthly Tanykina prices was obtained from Tanykina. Measured market prices were collected through monthly market price surveys in three markets (figure presents the three-market average). Estimated market price in a given week is the total sales value of milk supplied to the market that week, divided by the estimated quantity of milk sold in the market in that same week. The quantity of milk sold in the market is derived as the total quantity of milk sold (milk production net of milk consumption) minus the estimated quantity of milk sold to Tanykina (sales value of milk supplied to Tanykina divided by the Tanykina milk price). These data were measured as part of the Health and Financial Diaries data.

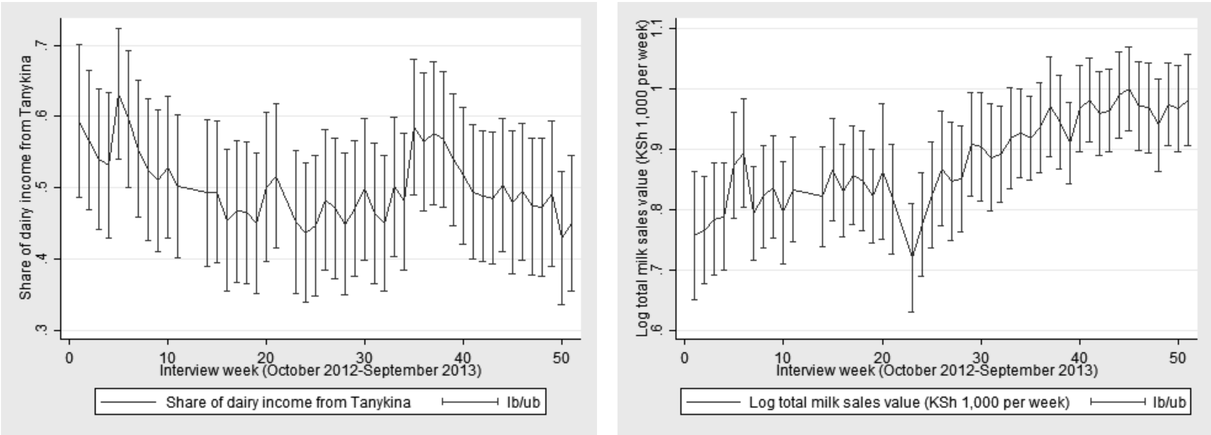
Figure 3.2 displays a number of variables related to the subsequent analyses over time. The top two graphs show for every week the sample mean along with a 95 percent confidence interval for households in the analysis sample, conditional on their selling milk that week. Our two main explanatory variables include the log of weekly milk production and the log of nondairy cash income (*CashInc*) as a ratio of nonfood expenditures (*NonFood*), that is, $\log(CashInc/NonFood) = \log(CashInc) - \log(NonFood)$. We use this log transformation to reduce the impact of outliers on our estimates.[11] Notice that milk production increases from the start of the diaries in October until 7 weeks later, which is early December. It then falls gradually over time, until Week 22, in March, when it starts increasing again until Week 44, in August. The log ratio of nondairy cash income and nonfood expenses does not correlate as much over time. It is particularly low in the weeks around Christmas (Weeks 11 and 14), and in the second week of May (Week 31), after which it starts to increase.

Figure 3.2 Main independent and dependent variables by week



[11] A log transformation of net income itself is unfeasible because net income is often negative.

Figure 3.2 Continued



Source: Data from the Health and Financial Diaries project in Kenya (Janssens et al., 2013).
Note KSh = Kenya shillings; lb = lower bound; ub = upper bound.

The next two graphs present the quantity of milk sold in a given week and the share of that milk sold to Tanykina. Households sell more milk in weeks with higher milk production, so the two variables are highly correlated. Further, households sell around 30 to 50 percent of their milk to Tanykina. This share is decreasing in the first period of the diaries, reaching its lowest point in the 10 weeks after Christmas. This is also the period when milk production is declining, while the income-expenditure ratio is fluctuating around its average value.

The last two graphs analyze milk *income*, and thereby combine the quantity of milk sold with the price received for that milk. These graphs describe the share of dairy income received from Tanykina (as opposed to income from other buyers in the local market) and the log of the total value of milk sold in a given week (in KSh 1,000), both conditional on the household's selling any milk that week. The average share of dairy income received from Tanykina fluctuates between 0.4 and 0.6. This is higher than the share of milk sold to Tanykina, indicating that Tanykina provides on average a higher price than buyers in the local market. We do, however, observe large variation between households; households may face heterogeneity in the local market price they can receive. Total dairy income drops in the week after Easter but is otherwise modestly increasing over time, due to both increasing price levels and higher milk production.

Table 3.2 summarizes statistics aggregated over the year for these and other variables used in the analyses. Columns (1)–(3) and (4)–(5) distinguish between households with and without variation over time in the proportion of dairy income received from Tanykina, respectively. Columns (6)–(7) present the difference in means between these two samples and the *p*-value from a *t*-test for differences in means, with standard errors clustered at the household level.

Average nondairy household income in cash is around KSh 2,000 (US $20) per week. Nonfood expenditures are on average KSh 2,500 (US $25), and weekly food expenditures amount to approximately one-fifth of this amount. In around one-quarter of all weeks, at least one household member faces a health problem, and in about one-third of all weeks, households have health insurance coverage. The average household produces 72 liters of milk per week, and around one-third of this milk is consumed by the household. The remainder is sold. We observe dairy sales in 85 percent of all

Conditional on selling milk in a given week, an average of 1.01 household members report selling milk. Total dairy income in these weeks is around KSh 1,500, filling the net gap between nondairy income and total nonfood and food expenditures. Even though Tanykina offers a higher milk price on average than the local market, only half of total dairy income is received from Tanykina; the remainder is received from other buyers. Only a few of these other buyers defer their payments. As a result, the share of dairy income to be received from selling milk on credit versus cash is closely correlated with the share of dairy income from selling milk to Tanykina versus other buyers.

Table 3.2 Summary statistics of time-varying characteristics

Variable	Variation in share of income from Tanykina: Mean (1)	s.e. (2)	within (3)	No variation in share of income: Mean (4)	s.e. (5)	Difference in means: Mean (6)	*p*-value (7)
Weekly nondairy cash income in 1,000 KSh	2.009	4.065	2.924	2.031	4.907	-0.052	0.9201
Weekly nonfood expenditures in 1,000 KSh	2.493	4.457	3.754	1.907	6.030	0.542	0.2866
Weekly food expenditures in 1,000 KSh	0.537	0.912	0.844	0.549	1.372	-0.014	0.8907
Prop. of weeks with health problem	0.263	0.440	0.391	0.271	0.445	-0.004	0.9239
Prop. of weeks with insurance coverage	0.344	0.475	0.245	0.390	0.488	-0.056	0.5379
Liters of milk produced per week	71.50	37.49	19.16	52.97	34.44	17.70	0.0083
Liters of milk consumed per week	18.16	7.022	4.772	14.46	6.384	3.587	0.0002
Share of milk that is consumed	0.316	0.196	0.137	0.339	0.186	-0.021	0.4649
Prop. of weeks that household buys milk	0.004	0.065	0.056	0.004	0.060	0.001	0.8815
Tanykina milk price (monthly price in KSh)	36.19	2.217	2.208	36.20	2.221	-0.039	0.3437
Market milk price (average for 3 buyers)	30.52	2.614	2.603	30.54	2.622	-0.050	0.2776
Prop. of weeks with dairy sales	0.847	0.360	0.276	0.697	0.460	0.150	0.0127
Conditional on selling milk that week...							
Number of respondents selling milk	1.007	0.084	0.081	1.001	0.032	0.006	0.0156
Total dairy income in 1,000 KSh	1.572	0.936	0.523	1.325	1.025	0.226	0.2414
- Share received by main dairy farmer	0.926	0.260	0.222	0.941	0.235	-0.016	0.5637
Dairy income from Tanykina in 1,000 KSh	0.871	0.917	0.551	1.031	1.216	-0.173	0.4482
- Share of total dairy income	0.503	0.413	0.232	0.629	0.483	-0.132	0.1881
Dairy income from other buyers in 1,000 KSh	0.700	0.609	0.346	0.295	0.426	0.399	0.0001
- Share of total dairy income	0.497	0.413	0.232	0.371	0.483	0.132	0.1881
Dairy income received on credit in 1,000 KSh	0.904	0.935	0.556	1.118	1.164	-0.228	0.3087
Dairy income received in cash in 1,000 KSh	0.667	0.606	0.335	0.207	0.400	0.453	0.0000
- Share received by main dairy farmer	0.950	0.218	0.169	1.000	0.000	-0.051	0.0010
Conditional on selling milk, share of milk...							
- sold to Tanykina*	0.300	0.309	0.227	0.395	0.419	-0.099	0.1235
- sold in local market*	0.329	0.290	0.169	0.228	0.312	0.104	0.1144
- consumed by the household	0.274	0.116	0.074	0.292	0.127	-0.016	0.4264
Number of households (total observations)	88	(3,997)		30	(1,381)		

Source: Data from the Health and Financial Diaries project in Kenya (Janssens et al., 2013).

Note: The sample excludes two households with high attrition and strong outliers. Further, in assessing whether there is variation over time in the share of income received from Tanykina, we omit weeks with high attrition due to Christmas (two weeks), Easter (one week) and the last fieldwork month (four weeks). The main dairy farmer is the household member reporting the highest value of dairy income received throughout the year. Estimated from dividing total sales value by the Tanykina and other buyers' milk prices, respectively. October milk prices were not collected and have been imputed by November milk prices. KSh = Kenyan shillings.

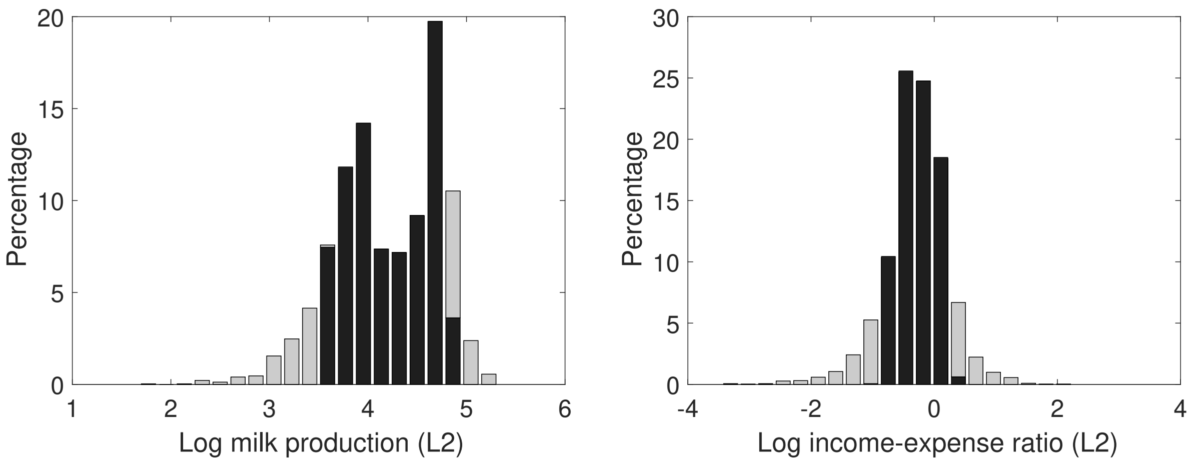
In weeks that households sell milk, they sell around 70 percent of their milk production; the remaining 30 percent is consumed by members of the household. We estimate that another 30 percent of production is sold to Tanykina, so that the remaining 40 percent is sold in the local market. This is an estimate, though, because interviewers did not record the quantity of milk sold to Tanykina; they reported total earnings from selling milk to Tanykina since the last interview. Although we can ascertain that the official monthly Tanykina price is accurate (it is constant across collection centers and does not vary within a month), farmers may not always have reported the gross value of their earnings from Tanykina. The possibility exists that they reported earnings net of deductions for milk transportation and shares. In that case, the 30 percent is underestimated, so that households sell more than 30 percent of their milk production to Tanykina and less than 40 percent to other buyers.

To shed more light on sales shares, note that conditional on selling, households receive on average KSh 700 from other buyers. In the weeks in which households sold milk, milk costs on average KSh 30.79 per liter in the local markets. This suggests that households sold on average 700/30.79 = 22.73 liters of milk per week. With an average milk production of 75.74 liters per week, these statistics suggest that at least 30 percent of milk production is sold in the local market.[12] Further, the same interviewer always visited the same households. Most likely, respondents will have reported either always gross earnings from Tanykina or always net earnings, depending on the interviewer, but not gross earnings in one week and net earnings in another week. Hence, household fixed effects will control for the potential measurement error in the share of milk sold to Tanykina.

Figure 3.3 presents the percentage histograms of our main independent and dependent variables. Panel indephist provides a histogram for the lagged log of milk production, *L2MilkProd*, and the lagged value of log nondairy income net of log nonfood expenditures, *L2NetInc*. The gray bars represent the upper and lower 10 percent. Given the scarcity of observations in the tails, nonparametric estimation in this range might be imprecise or even inaccurate. Thus, we will restrict our estimation of the nonparametric function across the 10 to 90 percent quantile range of *L2MilkProd* and *L2NetInc* using all observations in the sample.

Figure 3.3 Percentage histograms of main independent and dependent variables

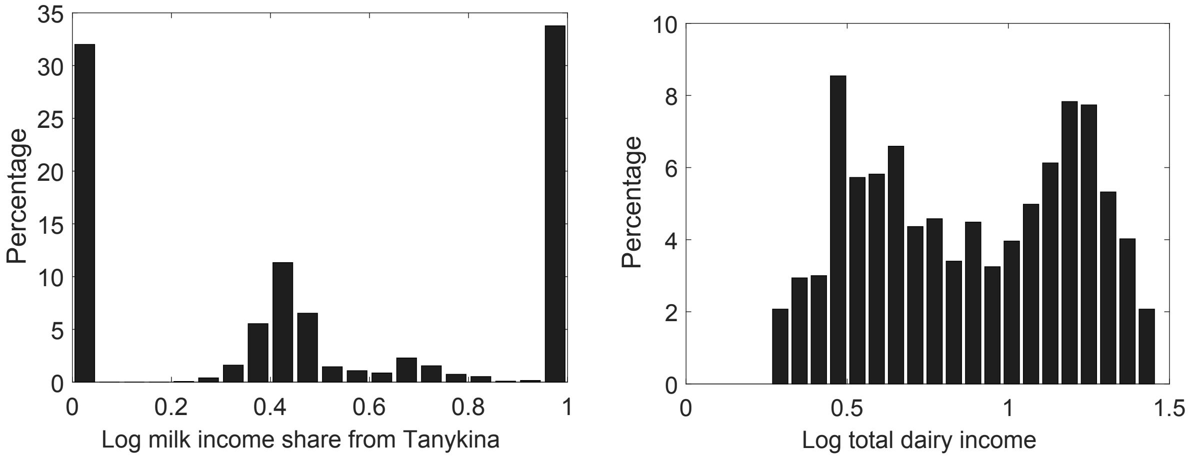
(a) Log milk production and log income-expense ratio (L2). Gray areas: upper and lower 10 percent



[12] Given that the average milk price is the price that a buyer would pay in the local market, the measured prices are likely an upper bound of the farmgate price. The 22.73 liters of milk per week is hence likely an underestimate of the total quantity of milk that a farmer actually sells in the local market.

Figure 3.3 Continued

(b) Histograms of the share of log dairy income earned from Tanykina and log total dairy income



Source: Data from the Health and Financial Diaries project in Kenya (Janssens et al., 2013).
Note: Figures include the analysis sample, including all weeks from the Health and Financial Diaries in which the household sold any milk, for households with variation over time in the share of dairy income received from Tanykina.

Panel (b) presents two of our main dependent variables, the share of log dairy income that is received from Tanykina (conditional on selling milk) and the log of total dairy income (conditional on selling milk). Households typically sell either no milk to Tanykina, or all of their milk; alternatively, they sell about half of their milk to Tanykina. It is very common for households to sell milk produced in the morning to Tanykina and milk produced in the afternoon in the local market. This explains the high number of households selling around half of their milk to the cooperative.

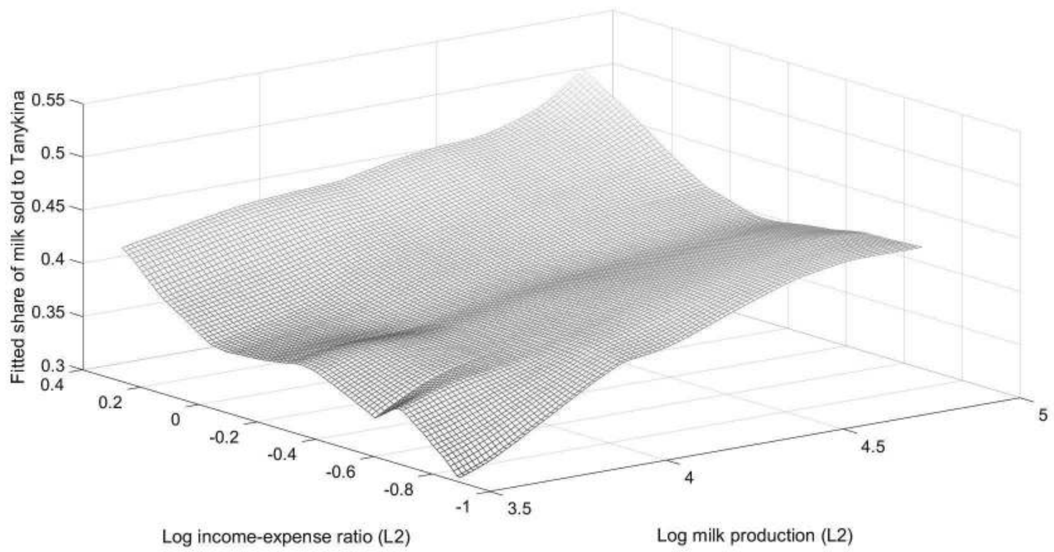
4. RESULTS

Semiparametric Estimates

This section presents the semiparametric regression results for our three main outcome variables: the share of marketed milk that is sold to Tanykina (*ShareMilkTan*), the share of dairy income from Tanykina (*ShareTan*), and total dairy income in a given week (*SalesAll*). Because the first two variables are defined only for weeks in which the household sells milk, the analyses in this section will focus on these weeks (about 85 percent of all weeks).

We first estimate Equation (3) for the share of milk sold to Tanykina and the share of dairy income from Tanykina, both conditional on selling milk. Equation (7) gives the piontwise local linear estimators for $m(\cdot)$ and the slopes. To visualize the joint effects of milk production and net income on where farmers sell their milk, Figure 4.1 presents a three-dimensional graph, plotting the fitted value of the share of milk sold to the cooperative ($\hat{m}(\cdot)$) against average milk production and net income in the past two weeks. These fitted values are rescaled to have the same mean as the dependent variable for ease of comparison between different specifications. The plot is derived using piontwise estimates of $m(\cdot)$ and its slopes at 100×100 evenly spaced points across the 10 to 90 percent quantile of $L2MilkProd$ and $L2NetInc$.

Figure 4.1 Fitted share of milk sold to Tanykina against lagged net income and milk production

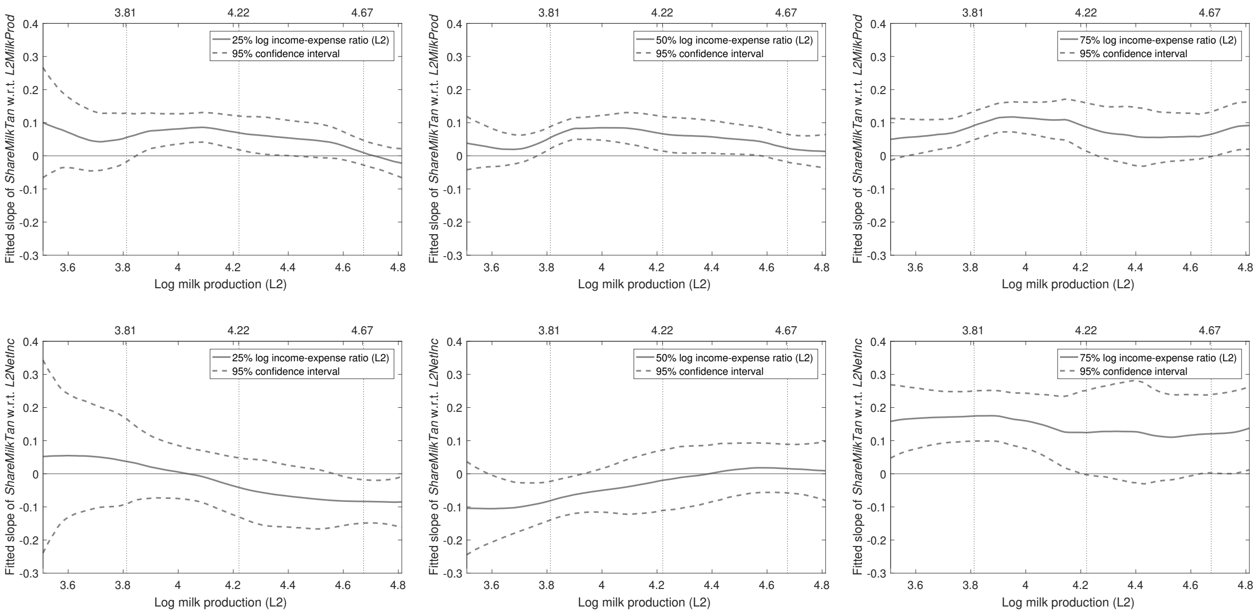


Source: Data from the Health and Financial Diaries project in Kenya (Janssens et al., 2013).
Note: *L2MilkProd* = log milk production in liters; *L2NetInc* = log nondairy cash income minus log nonfood expenditures; *ShareMilkTan* = share of marketed milk sold to Tanykina; w.r.t. = with respect to.

In Figure 4.1, the fitted share of milk that is sold to Tanykina is at its lowest point in weeks that follow a period of both low net nondairy income and low milk production. In weeks that follow a period of either higher income or higher milk production, households sell a relatively larger share of milk to Tanykina. They do not respond linearly to changes in net income; the share of milk sold to Tanykina increases in net income mainly at higher levels of net income. This finding is consistent with the notion of a threshold effect; cash on hand needs to be sufficiently large for it to have an effect on farmers' decision of where to sell their milk.

Figure 4.1 presents fitted *levels* of our main outcome variables. In order to give a closer examination of the individual effect of each covariate, we analyze how the share of milk sold to Tanykina responds to our explanatory variables by presenting the fitted *slopes* in Figure 4.2. Figure 4.2 presents these slopes with respect to lagged milk production in the top panel and with respect to lagged net income in the bottom panel, together with 95 percent confidence intervals. The figure evaluates these slopes at different levels of lagged milk production (varied on the horizontal axis) and lagged net income (at the 25th, 50th and 75th percentiles in the left, middle and right figures, respectively). To make the plot more informative, we also mark the 25th, 50th and 75th percentiles of milk production on the top axis and draw three vertical dotted lines to indicate these quantiles.

Figure 4.2 Fitted slopes of the share of milk sold to Tanykina and 95 percent confidence intervals



Source: Data from the Health and Financial Diaries project in Kenya (Janssens et al., 2013).
Note: *L2MilkProd* = log milk production in liters; *L2NetInc* = log nondairy cash income minus log nonfood expenditures; *ShareMilkTan* = share of marketed milk sold to Tanykina; w.r.t. = with respect to.

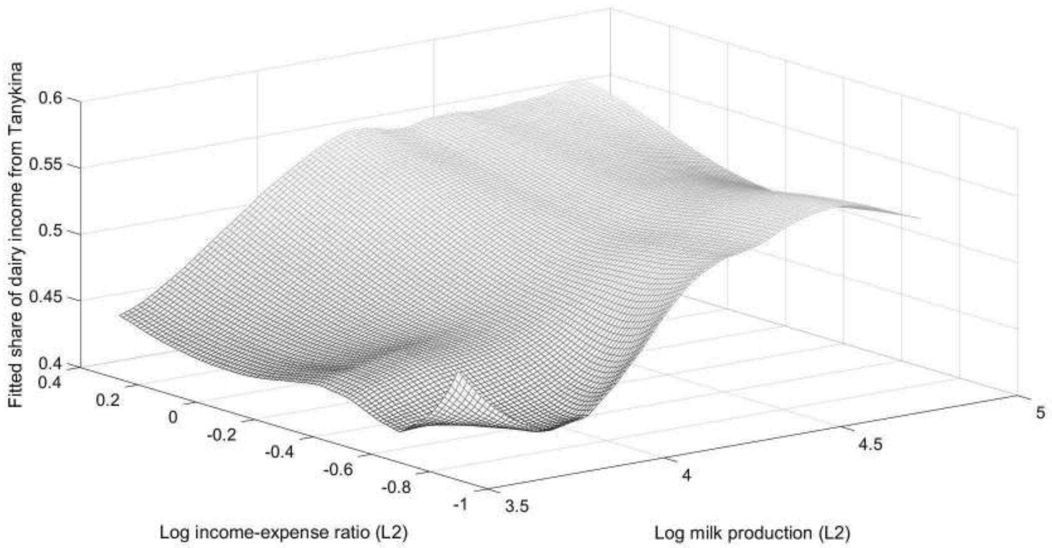
In the top panel of Figure 4.2, we observe a significantly positive slope at all three levels of the log income-expense ratio, mainly when milk production is just below its median level. The share of milk sold to Tanykina is also increasing in milk production at lower levels of milk production, but only when net income is around the median level or higher. Around the median level of milk production and at the median net income level (that is, the middle graph in the top panel), a 1 percent increase in milk production increases the share of milk sold to Tanykina by approximately 5 percentage points.

The bottom panel of Figure 4.2 presents fitted slopes for the share of milk sold to Tanykina with respect to average nondairy income net of nonfood expenditures in the past two weeks. The slope is close to 0 and not statistically significant when net income is at the 25th or 50th percentile, irrespective of the lagged milk production level. We do, however, find a positive effect of net income at the 75th percentile of net income, with more precise estimates at below-median levels of milk production. In weeks with net income at the 75th percentile and past milk production at the 25th percentile, a 1 percent increase in the income-expenditure ratio raises the share of milk sold to Tanykina by nearly 20 percentage points.

In a similar way, Figures 4.3 and 4.4 present fitted levels and slopes, respectively, of the share of dairy income received from Tanykina, taking into account not only changes in the share of milk sold to Tanykina but also differences in the price offered by Tanykina versus other buyers. A simple comparison between Figure 4.2 (share of milk sold to Tanykina) and 4.4 (share of dairy income from

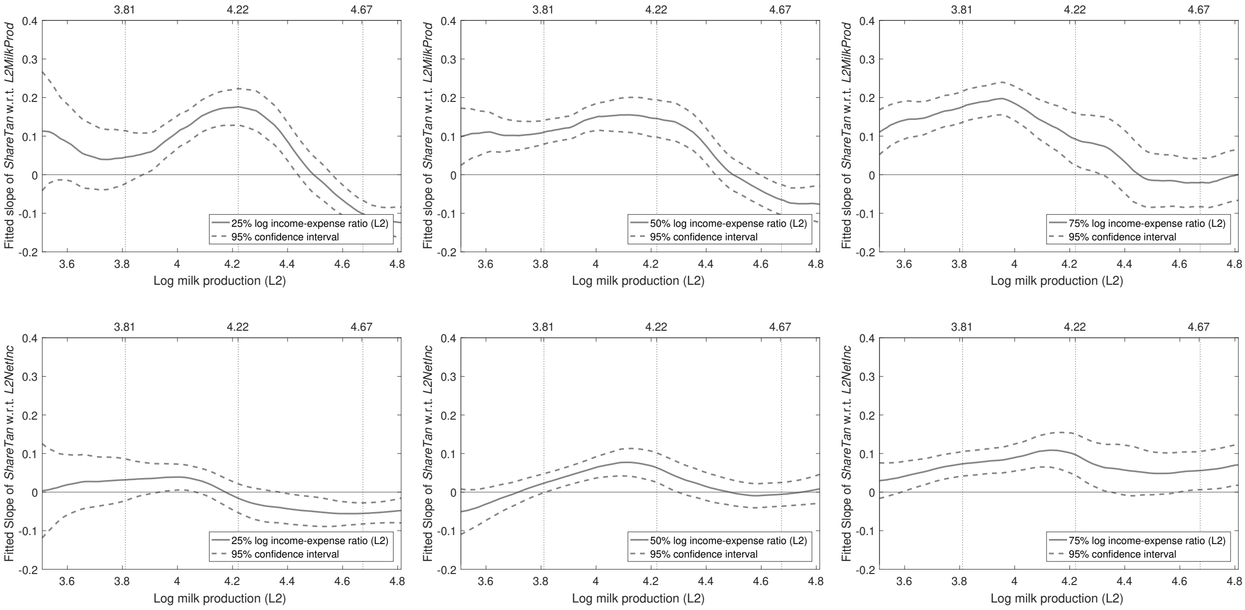
In a similar way, Figures 4.3 and 4.4 present fitted levels and slopes, respectively, of the share of dairy income received from Tanykina, taking into account not only changes in the share of milk sold to Tanykina but also differences in the price offered by Tanykina versus other buyers. A simple comparison between Figure 4.2 (share of milk sold to Tanykina) and 4.4 (share of dairy income from Tanykina) reveals three key findings. First, milk production has a larger and more significant effect on the share of income from Tanykina than it does on milk sold, in fact double the effect at median levels of net income and milk production, suggesting that at such income and production levels, farmers receive a relatively low milk prices in the local market. If the relative price were constant, the effects of dairy production and nondairy income on the share of income from the cooperative would be similar to the effects on the share of milk sold to the cooperative. Second, we observe this effect only at lower levels of milk production; at higher levels of milk production, the effect on the share of dairy income from Tanykina is *lower* than the effect on the share of milk sold to Tanykina. Third, the positive effect of net income at below-median levels of milk production, shown in the lower right-hand panel of Figure 4.4 for the 75th percentile of net income, is only half the size of the effect on the share of milk sold to Tanykina (same panel of Figure 4.2). This finding indicates that at such levels of milk production and net income, farmers receive a relatively high price for milk sold in the local market.

Figure 4.3 Fitted share of dairy income from Tanykina against lagged net income and milk production



Source: Data from the Health and Financial Diaries project in Kenya (Janssens et al., 2013).

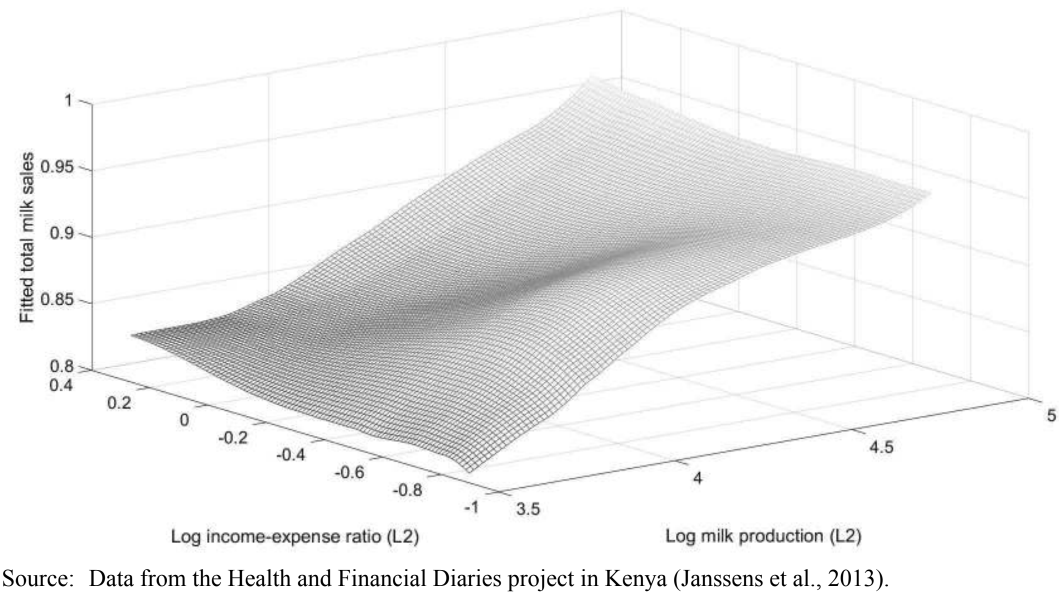
Figure 4.4 Fitted slopes of the share of dairy income from Tanykina and 95 percent confidence intervals



Source: Data from the Health and Financial Diaries project in Kenya (Janssens et al., 2013).
Note: *L2MilkProd* = log milk production in liters; *L2NetInc* = log nondairy cash income minus log nonfood expenditures; *ShareMilkTan* = share of marketed milk sold to Tanykina; w.r.t. = with respect to.

Figure 4.5 analyzes whether these effects of cash on hand on milk marketing decisions have implications for households' dairy income levels. The figure estimates Equation (3) using a household's weekly dairy income as the dependent variable. It presents fitted levels of dairy income against lagged milk production and the lagged income-expense ratio. Figure 4.6 shows the slopes of fitted dairy income associated with these two variables.

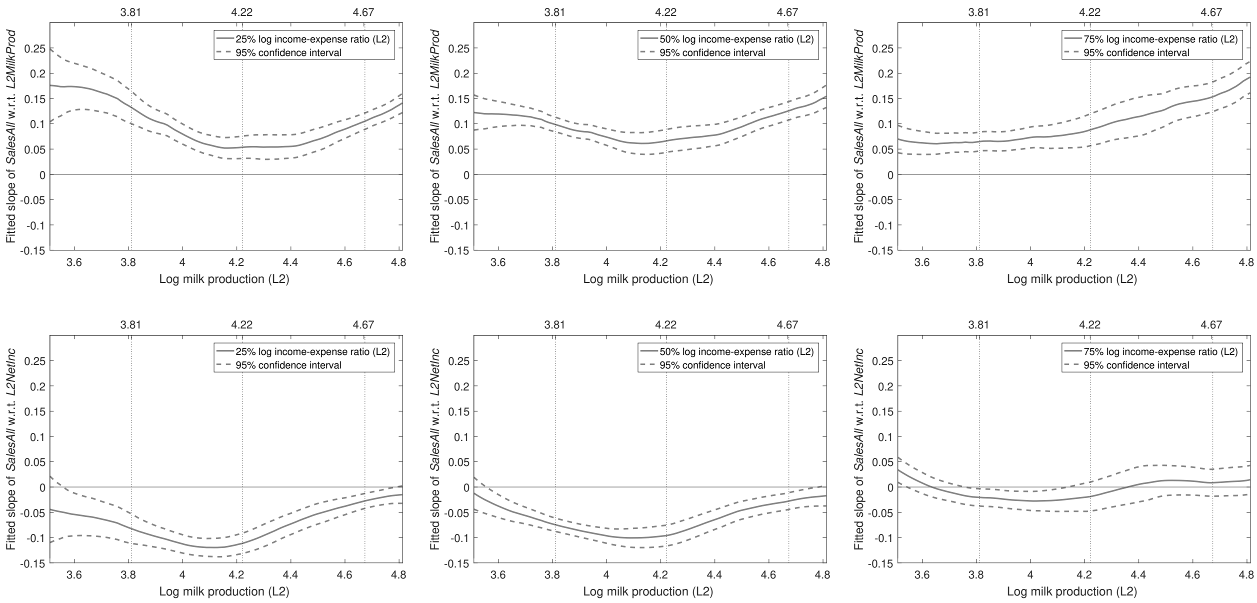
Figure 4.5 Fitted total dairy income against lagged net income and milk production



Source: Data from the Health and Financial Diaries project in Kenya (Janssens et al., 2013).

Total income is increasing in lagged milk production in Figure 4.6. This is not necessarily driven by a substitution away from the local market to Tanykina as milk production increases. We will later show that milk consumption is decreasing in lagged milk production. The positive effect of lagged milk production on total dairy income may hence be related to an effect on total quantity sold instead of an effect on where the milk is sold.[13]

Figure 4.6 Fitted income share from Tanykina against lagged net income and milk production



Source: Data from the Health and Financial Diaries project in Kenya (Janssens et al., 2013).
Note: *L2MilkProd* = log milk production in liters; *L2NetInc* = log nondairy cash income minus log nonfood expenditures; *ShareMilkTan* = share of marketed milk sold to Tanykina; w.r.t. = with respect to.

The bottom panel of Figure 4.6 presents the slope of total dairy income with respect to the lagged income-expense ratio. We find a small negative effect of net income on total dairy income in weeks with below-median levels of net income. We do not observe this effect in weeks when a higher level of cash on hand allows households to sell more milk to Tanykina. This suggests that when households sell more milk in the local market to cope with a reduction in net income, they do not receive a lower price than they would have received from Tanykina. Recall from Figure 3.1 that, on average, the local market offers lower prices than Tanykina. Hence, buyers in the local market are offering relatively higher prices in weeks when reductions in net income induce farmers to sell their milk in the local market.

Table 4.1 provides the estimates for variables that are included in the linear part of the semiparametric model. Lagged food expenditures are not related significantly to the decision of where farmers sell milk, and not to dairy income either. Health shocks are significantly related to milk marketing decisions. As shown in Column (2), the share of milk sold to Tanykina is a significant 7.1 percentage points lower just after household members report health symptoms than in weeks without health symptoms. It is very likely that when someone is ill, households sell more milk in the local market in order to obtain cash to cope with medical expenditures, for instance to repay health providers. Consistent with this interpretation, we do not observe a negative effect of past health symptoms on marketing decisions for households with health insurance coverage; the interaction term of health insurance coverage and health symptoms is significantly positive and nearly outweighs the

[13]Note that we include current milk production as a control variable. Hence, serial correlation in milk production cannot account for this result.

negative effect of health symptoms. Health insurance coverage may well improve compliance with informal collective marketing arrangements. Past health symptoms have smaller coefficients in specifications for the share of dairy income from Tanykina, suggesting that the local market price is relatively high for farmers in weeks when health symptoms induce them to sell in the local market. In weeks without health symptoms, variation in insurance coverage over time does not affect where farmers sell their milk, but insurance coverage is related to higher dairy income, as shown in Columns (5) and (6). Current milk production and consumption are not related significantly to the share of milk sold to Tanykina. However, the share of dairy income from Tanykina and total dairy income are increasing in current milk production and decreasing in milk consumption.

Table 4.1 Factors affecting where the household sells the milk and total dairy income

	Share of milk sold to Tanykina		**Share of dairy income from Tanykina**		**Log dairy income from all buyers**	
Variable	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Lagged variables (average over past two weeks)						
Log liters of milk production	N/A	N/A	N/A	N/A	N/A	N/A
Log income-expense ratio	N/A	N/A	N/A	N/A	N/A	N/A
... *X* Log liters of milk production	N/A	N/A	N/A	N/A	N/A	N/A
Log food expenditures in 1,000 KSh	0.039	0.037	0.022	0.008	0.023	-0.000
	(0.042)	(0.043)	(0.026)	(0.026)	(0.019)	(0.015)
HH member has health symptoms	-0.072***	-0.071***	-0.052**	-0.057**	-0.018	-0.028***
	(0.023)	(0.023)	(0.025)	(0.025)	(0.011)	(0.010)
HH has insurance coverage	-0.013	-0.016	0.007	-0.006	0.050**	0.025*
	(0.033)	(0.033)	(0.028)	(0.028)	(0.020)	(0.015)
... *X* HH member has health symptoms	0.053*	0.054*	0.036	0.046	-0.005	0.014
	(0.031)	(0.032)	(0.031)	(0.031)	(0.019)	(0.017)
Other variables						
Current log liters of milk production		-0.065		0.172***		0.379***
		(0.055)		(0.023)		(0.020)
Current log liters of milk consumed		0.086		-0.042*		-0.129***
		(0.057)		(0.021)		(0.020)
District-month effects	Yes	Yes	Yes	Yes	Yes	Yes
R-squared within households	0.153	0.157	0.125	0.152	0.292	0.501
Mean dependent variable	0.410	0.410	0.502	0.502	0.888	0.888
Number of observations	3,231	3,231	3,231	3,231	3,231	3,231
Number of households	88	88	88	88	88	88

Source: Data from the Health and Financial Diaries project in Kenya (Janssens et al., 2013).
Note: Analysis sample includes only households with variation in the share of milk sold to Tanykina. The log of variable x is proxied by $\log(x+1)$. We report R^2 in the fixed-effect and semiparametric regressions by R^2 within, that is, the squared correlation coefficient between the demeaned y (minus household mean across time) and fitted value of demeaned y, $\tilde{y}_{it}$ and $\hat{\tilde{y}}_{it}$. In particular, for the semiparametric regression, we have $\tilde{y}_{it} \equiv y_{it} - \bar{y}_i$ and $\hat{\tilde{y}}_{it} \equiv (x_{it} - \bar{x}_i)'\hat{\beta} + \hat{m}(z_{it}) - \bar{\hat{m}}_i$, where $\bar{y}_i \equiv 1/T\sum_{t=1}^{T} y_{it}$, $\bar{x}_i \equiv 1/T\sum_{t=1}^{T} x_{it}$, and $\bar{\hat{m}}_i \equiv 1/T\sum_{t=1}^{T} \hat{m}(z_{it})$. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. HH = household; KSh = Kenyan shillings.

In sum, our findings suggest that farmers sell relatively more milk in the local market in weeks that follow a period of lower milk production, lower nondairy income, or higher nonfood expenditures, that is, weeks following a period in which households were able to accumulate less cash on hand. The effects of nondairy income exist only at relatively high income levels when milk production is low, providing evidence of threshold effects and nonlinearities. Further, to the extent that farmers switch away from Tanykina to other buyers in weeks when they have less cash on hand, they do not give up substantial amounts of income by doing so. Milk prices in the local market are on average lower than those offered by Tanykina, but in weeks when households are induced to sell more in the local market, that is, in weeks that follow a period of low milk production and low levels of net income, they may be able to bargain for better prices. This provides a potential explanation for why farmers with lower levels of cash from both past dairy sales and current nondairy income are not induced to sell in the local market; their low bargaining power may push prices down to a point where it is more beneficial to use alternative ways to obtain cash.

Fully Linear Estimates

Next, we compare the findings of the semiparametric estimates with the results from the fully linear model described in Section 3. Table 4.2 estimates Equation (2), which replaces the nonparametric part, $m(z_{it})$, with linear terms for lagged milk production and log income net of log expenses, as well as an interaction of the two. As in the case of the semiparametric results, we include as dependent variables the share of milk sold to Tanykina in Columns (1)–(2), the share of dairy income from Tanykina in Columns (3)–(4), and log total dairy income in Columns (5)–(6). Even-numbered columns control for current milk production and consumption. For ease of comparison with the semiparametric results, Panel A centers lagged milk production and the income-expense ratio at their 25th percentile levels, Panel B centers these variables at their median levels, that is, the 50th percentile, and Panel C at the 75th percentile. Although estimates of the interaction term do not depend on where the explanatory variables are centered, the main effects are sensitive to where the other variable is evaluated. The semiparametric estimates are not subject to this sensitivity because they evaluate the effect of the main variables locally.

Using semiparametric estimates, we conclude the following:

(SP1) At all levels of lagged net income, the share of milk sold to Tanykina and the share of dairy income from Tanykina are both increasing in lagged milk production at below-median levels.

(SP2) This effect is larger and more significant for the share of income, with a more obvious deceasing return to scale.

(SP3) Lagged net income affects these two outcome variables, but only at the 75th percentile and only in weeks with below-median milk production.

(SP4) This effect is larger for the share of milk sold to Tanykina.

(SP5) In weeks with more milk production and less net income, households are more likely to have a higher total dairy income.

The fully linear estimates in Table 4.2 replicate a number of these findings but yield opposite conclusions regarding others. In Panels A and B of Columns (2) and (4), the effect of lagged milk production is positive and statistically significant (with one exception), in line with Result (SP_1). Nonetheless, the estimated coefficients are smaller than the estimated slopes in the semiparametric framework.

Further, the effect of lagged milk production is decreasing in the level of lagged milk production; but this decreasing return to scale is not much stronger for the share of dairy income from Tanykina, as is the case in the semiparametric model. Hence, the linear estimates are only partially consistent with Result (SP_2).

Regarding Result (SP_3), the positive effect of net income is found to be significant only when milk production is around or below its median level. This finding is inconsistent with the semiparametric finding that net income from nondairy activities has an effect only in weeks with high net income.

Table 4.2 Estimates of the fully linear model for milk market decisions

Variable	Share of milk sold to Tanykina		Share of dairy income from Tanykina		Log dairy income from all buyers	
	(1)	(2)	(3)	(4)	(5)	(6)
Lagged variables (average over past two weeks)						
Panel A. Centered at 25th percentile						
Log liters of milk production	0.001	0.056*	0.112***	0.045**	0.252***	0.076***
	(0.028)	(0.031)	(0.018)	(0.020)	(0.012)	(0.011)
Log income-expense ratio	0.096***	0.095***	0.035*	0.036**	0.022*	0.025**
	(0.028)	(0.028)	(0.018)	(0.018)	(0.012)	(0.010)
... *X* Log liters of milk production	-0.079***	-0.079***	-0.033*	-0.033*	-0.031***	-0.032***
	(0.028)	(0.028)	(0.018)	(0.018)	(0.012)	(0.010)
Panel B. Centered at 50th percentile						
Log liters of milk production	-0.021	0.033	0.102***	0.036*	0.243***	0.066***
	(0.026)	(0.029)	(0.017)	(0.019)	(0.011)	(0.011)
Log income-expense ratio	0.066***	0.065***	0.023*	0.023*	0.011	0.013*
	(0.020)	(0.020)	(0.013)	(0.013)	(0.009)	(0.007)
... *X* Log liters of milk production	-0.079***	-0.079***	-0.033*	-0.033*	-0.031***	-0.032***
	(0.028)	(0.028)	(0.018)	(0.018)	(0.012)	(0.010)
Panel C. Centered at 75th percentile						
Log liters of milk production	-0.046*	0.009	0.092***	0.025	0.233***	0.056***
	(0.026)	(0.030)	(0.017)	(0.019)	(0.011)	(0.011)
Log income-expense ratio	0.018	0.017	0.003	0.003	-0.009	-0.006
	(0.017)	(0.017)	(0.011)	(0.011)	(0.007)	(0.006)
... *X* Log liters of milk production	-0.079***	-0.079***	-0.033*	-0.033*	-0.031***	-0.032***
	(0.028)	(0.028)	(0.018)	(0.018)	(0.012)	(0.010)
Other variables						
Log food expenditures in 1,000 KSh	0.034	0.037	0.029	0.025	0.021	0.011
	(0.036)	(0.036)	(0.023)	(0.023)	(0.016)	(0.013)
HH member has health symptoms	-0.085***	-0.084***	-0.068***	-0.069***	-0.024**	-0.027***
	(0.027)	(0.027)	(0.018)	(0.017)	(0.012)	(0.010)
HH has insurance coverage	-0.026	-0.022	-0.004	-0.010	0.046***	0.031***
	(0.028)	(0.028)	(0.018)	(0.018)	(0.012)	(0.010)
... *X* HH member has health symptoms	0.056**	0.051**	0.058**	0.061**	0.009	0.017
	(0.026)	(0.026)	(0.028)	(0.028)	(0.019)	(0.016)
District-month effects	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Milk production and consumption	No	Yes	No	Yes	No	Yes
R-squared within households	0.062	0.068	0.104	0.121	0.274	0.487
Mean dependent variable	0.410	0.410	0.502	0.502	0.888	0.888
Number of observations	3,231	3,231	3,231	3,231	3,231	3,231
Number of households	88	88	88	88	88	88

Source: Data from the Health and Financial Diaries project in Kenya (Janssens et al., 2013).
Note: Analysis sample includes only households with variation in the share of milk sold to Tanykina. The log of variable x is proxied by $\log(x+1)$. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Controls: Average over past two weeks of log food expenditures in 1,000 KSh, HH member has health symptom, HH has insurance coverage, and the interaction of the last two variables. HH = household; KSh = Kenyan shillings.

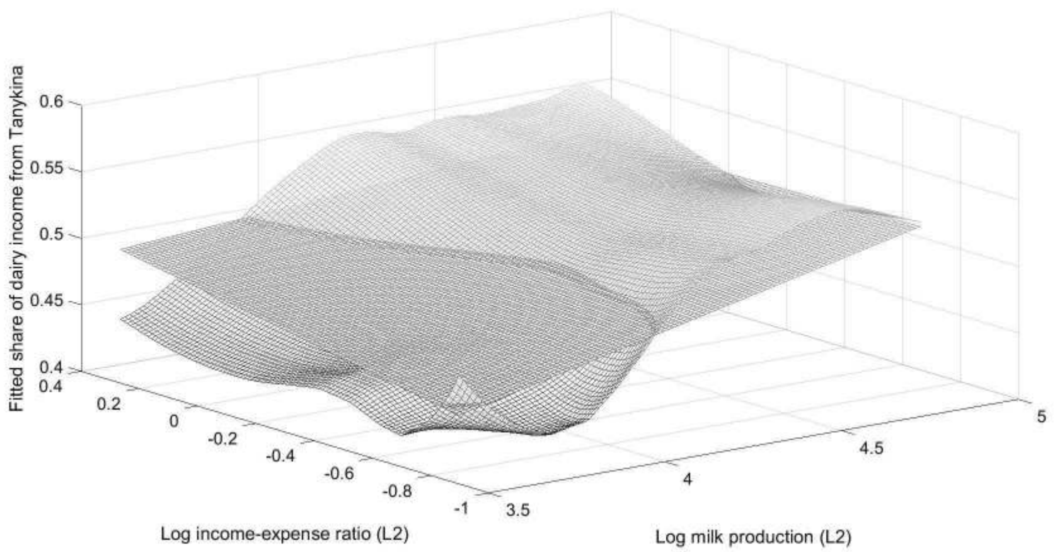
The linear model also cannot fully replicate Result (SP_4). In weeks that follow a period of higher milk production, total dairy income tends to increase, but the negative interaction term implies that this effect is diminishing in net income. The semiparametric estimates found that the positive effect of milk production is strongest when milk production and net income are either both low or both high.

Finally, Result (SP_5)—the negative effect of net income on total dairy income at below-median levels of milk production and net income—is not replicated either. As shown in Column (6), a 1 percent increase in the income-expense ratio increases dairy income by 2.5 percent when both variables are evaluated at the 25th percentile (Panel A), and when evaluated at the 50th percentile, a 1 percent increase in net income increases dairy income by 1.3 percent.

Interpretation of these differences between the linear and semiparametric results is straightforward but important. The linear model allows the income-expense ratio effect to vary only in log milk production, not in its own level, and only linearly. In other words, the increase of dairy income by 2.5 percent is assumed to be the same across the entire range of the income-expense ratio, and a 1 percent increase in log milk production is assumed to decrease this effect by a constant 3.2 percent, leading to an almost 0 effect around the median of log milk production and a large negative effect when log milk production is high. The nonparametric estimator allows this effect to vary in both production and net income in a nonlinear manner. As a result, it is largest around the median of log milk production when the income-expense ratio is not high, and smallest when milk production is at the two tails of its distribution or the income-expense ratio is high. The linear specification averages out these local effects, biasing the estimates and inference.

To illustrate the main difference between the semiparametric and fully linear estimates, Figure 4.7 stacks graphs for the fitted share of dairy income from Tanykina using the semiparametric estimates and the linear coefficient estimates of lagged milk production, nondairy net income, and their interaction term. Both graphs are rescaled to have the same mean for ease of comparison. In stark contrast to the flat plane estimated using the linear model, the nonparametric estimation captures the varying effect of lagged milk production by itself. In particular, after weeks with low milk production, an increase in milk production raises the share of dairy income from Tanykina substantially more than in weeks after a period of high milk production. The linear estimate can be interpreted as an average of the nonparametric estimates across both explanatory variables, leaving out more detailed information from local effects.

Figure 4.7 Fitted income share from Tanykina against lagged net income and milk production



Source: Data from the Health and Financial Diaries project in Kenya (Janssens et al., 2013).

Finally, the switch from a parametric to a semiparametric setting improves the explained variation in our outcome variable within households by comparing the same columns from Tables 4.1 and 4.2. In addition, the fully linear model does not allow for the threshold effects that we observe, and does not shed light on where in the distribution we find effects of cash on hand.

5. CONCLUSION

This paper analyzed to what extent cash on hand affects dairy farmers' compliance with collective marketing arrangements. While cooperatives typically defer payments, buyers in local markets are used to paying farmers in cash. As a result, a cash constrained farmer may decide to sell milk in the local market even when the local market offers a lower price than the cooperative. Our conceptual framework formalizes this hypothesis, and further shows that the share of milk sold to the cooperative is a nonlinear function of cash on hand, requiring cash on hand to reach certain thresholds before it affects farmers' decisions on where to sell their milk.

We then test this hypothesis using high-frequency, high-detail panel data for a sample of dairy farmers in western Kenya, whereby we rely on semiparametric specifications. We proxy the extent to which a household is cash constrained by measuring levels of lagged milk production—as a proxy for cash inflows from the cooperative—and nondairy income net of nonfood expenditures—as a proxy for net inflows of cash from income-generating activities other than dairy farming. We estimate a semiparametric model that includes household fixed effects. In this way, we explore how, within the same household, time variation in cash on hand influences our main outcome variables, that is, the share of milk sold to the cooperative, the share of dairy income from the cooperative, and the total dairy income. The semiparametric model provides a richer description by evaluating effects locally, which is appropriate given that we expect our outcome variables to be a nonlinear function of cash on hand, characterized by threshold effects.

Using this approach, we find that the share of milk sold to—and the share of dairy income from—the cooperative are increasing in lagged milk production at below-median levels; this holds true for all levels of lagged net income. Further, this effect is larger and more significant for the share of income from the cooperative, with a more obviously deceasing return to scale. Thus, after weeks in which households are able to sell more milk, increasing their net cash flow from dairy farming, they are less likely to sell outside the cooperative. Lagged net income affects these two outcome variables but only at the 75th percentile—suggestive of threshold effects—and only in weeks with below-median milk production. This effect is larger for the share of milk sold to Tanykina. Finally, after weeks with *more* milk production and *less* net income, households report earning more dairy income, even when we condition on current milk production and consumption, so that the outcome variable captures variation in milk prices, not in the quantity of milk sold. Hence, farmers receive, on average, higher prices after weeks with increased cash flows from dairy farming and weeks with less cash on hand from other sources.

We also find evidence of households' coping with health shocks by selling more milk to buyers in the local market, who pay in cash, than to the cooperative, whose payments are deferred. However, this coping strategy is not observed when households are covered by health insurance. Health expenditures and income losses associated with health shocks will make households more cash constrained, in particular when they are not covered by health insurance and have limited access to other coping strategies. In this light, these findings provide further evidence of the implications that cash on hand can have on the decision of where to sell milk. Further, because health insurance shields households from expenditures and is not associated with increased side selling, our findings highlight linkages between agriculture and health—that is, the potential of health insurance to strengthen collective marketing arrangements. To the best of our knowledge, such linkages have not been documented in previous literature.

In conclusion, our findings indicate that cash constraints influence where farmers sell their milk. When relatively cash constrained, they sell larger quantities of milk in the local market, where they are paid cash immediately, while higher levels of cash on hand allow them to deliver their milk to the cooperative and save for future expenditures. The cooperative offers a price that is, on average, higher than the price for milk sold in the local market. As a result, we find that a reduction in past milk production—associated with lower payments from the cooperative in the current period—lowers farmers' income, which may reinforce their preference for selling milk in the local market. Access to financial instruments that can help relax cash constraints, for instance health insurance, can thereby improve farmers' loyalty to the cooperative agreement and have positive externalities beyond improving farmers' health.

APPENDIX A: THEORETICAL PREDICTIONS FOR THE EFFECT OF MILK PRODUCTION AND NON-DAIRY INCOME ON SIDE-SELLING

Appendix A gives a detailed analysis of the effect of changes in predetermined milk production, m_{t-1}, and nondairy net income, y_t, on the amount of milk sold in the local market, s_t.

The first-order derivative of lifetime utility (1) in s_t is

$$p_t u'(y_t + p_t s_t + m_{t-1} - s_{t-1}) - \beta u'(y_{t+1} + p_{t+1} s_{t+1} + m_t - s_t). \tag{A.1}$$

In order to analyze the effect of changes in m_{t-1} and y_t in this framework with an infinite horizon, we first suppose that income, prices, and milk production are stable over time, which we denote by $\bar{y}$, $\bar{p}$, and $\bar{m}$, respectively. In that case, the quantity sold in the local spot market should also be stable. We denote this quantity $\bar{s}$. Then the first-order derivative (A.1) becomes

$$\bar{p}u'(\bar{y} + \bar{p}s_t + \bar{m} - \bar{s}) - \beta u'(\bar{y} + \bar{p}\bar{s} + \bar{m} - s_t). \tag{A.2}$$

Lemma 1 illustrates where farmers sell their milk in equilibrium under various conditions. Assume that the optimal s_t is determined only by the contemporaneous first-order condition (A.2), taking s_{t+1} at the equilibrium level, $\bar{s}$.

Lemma 1. *Consider the utility maximization problem given in* (1). *An equilibrium exists in which (1) if $\bar{p} < \beta$, the farmer sells all milk to the cooperative, that is, $\bar{s} = 0$; (2) if $\bar{p} > \beta$, the farmer sells all milk in the local spot market, that is, $\bar{s} = \bar{m}$; (3) if $\bar{p} = \beta$, the farmer is indifferent about where to sell, that is, $\bar{s} \in (0, \bar{m})$.*

Proof. Suppose the farmer always sells to the cooperative, $\bar{s} = 0$. In a given period, the first-order derivative evaluated at $s_t = 0$ reduces to

$$\bar{p}u'(\bar{y} + \bar{m}) - \beta u'(\bar{y} + \bar{m}) < 0, \qquad \forall\, \bar{p} < \beta, \tag{A.3}$$

and the farmer will not sell any milk in the local spot market. Likewise, suppose the farmer always sells in the local spot market, $\bar{s} = \bar{m}$. In a given period, the first derivative of utility evaluated at $s_t = \bar{m}$ is

$$\bar{p}u'(\bar{y} + \bar{p}\bar{m}) - \beta u'(\bar{y} + \bar{p}\bar{m}) > 0, \qquad \forall\, \bar{p} > \beta, \tag{A.4}$$

and the farmer will not sell any milk to the cooperative. Thus, with stable income, prices, and milk production, farmers will deliver their milk to the cooperative if $\bar{p} < \beta$hen the market price is relatively low and future payments are not discounted heavily, while they sell their milk in the local spot market if $\bar{p} > \beta$, when discounting is stronger and the market offers a relatively higher price. Finally, if $\bar{p} = \beta$, farmers are indifferent as to where to sell, but the equilibrium is unstable since slight $\bar{s}$ will fluctuate between 0 and $\bar{m}$ due to changes in y_t or m_{t-1}, as will be shown later, in the section. □

The remainder provides a qualitative analysis of how the quantity of milk sold in local spot markets (s_t) varies in the net nondairy income and predetermined milk production (y_t, m_{t-1}). We consider the three cases in equilibrium: one in which farmers sold all milk to the cooperative, and one in which they sold all milk previously in the local market, and a third in which farmers are indifferent as to where to sell their milk.

Farmers Sold All Milk to the Cooperative, $\bar{s} = 0$.

Consider the case in which $\bar{p} < \beta$. Lemma 2 analyzes how farmers respond to changes in (y_t, m_{t-1}) when they sold all their milk to the cooperative in the previous period. Assume that the potential compensating effect of s_{t-1} caused by changes in m_{t-1} is smaller than changes in m_{t-1} by itself.

Lemma 2. *Consider the utility maximization problem given in* (1) *in the equilibrium when* $\bar{p} < \beta$ *such that* $\bar{s} = 0$*. Then, (1) if there is an increase in current income,* $y_t > \bar{y}$*; an increase in past milk production,* $m_{t-1} > \bar{m}$*; or both, then* $s_t = 0$*; (2) if there is a decrease in* y_t*,* m_{t-1}*, or both, such that* $u'(y_t + m_{t-1})/u'(\bar{y} + \bar{m}) > \beta/\bar{p}$*, then* $s_t > 0$*. Further, if the response* s_t *satisfies* $u'(\bar{y} + \bar{m} - s_t)/u'(\bar{y} + \bar{m}) > \beta/\bar{p}$*, then also* $s_{t+1} > 0$*.*

Proof. When there is an increase in y_t, m_{t-1}, or both, it is straightforward that s_t remains at 0 because the negative first-order derivative (A.3) now becomes even more enforced due to (1) an increase in y_t, m_{t-1}, or both; (2) concavity of the utility function; and (3) the assumption that changes in s_{t-1} caused by m_{t-1} is not enough to compensate changes of m_{t-1} in itself, which is generally true.[14]

Now suppose that income in the current period t reduces to $y_t < \bar{y}$, milk production at time $t-1$ reduces to $m_{t-1} < \bar{m}$, or both. In that case, the marginal utility of selling milk in the spot market at time t, evaluated at $s_t = 0$ (so that the next period, the farmer earns $\bar{y} + \bar{m}$, and will again not sell any milk in the local spot market, $s_{t+1} = 0$), is[15]

$$\bar{p}u'(y_t + m_{t-1}) - \beta u'(\bar{y} + \bar{m}) > 0, \quad \textit{iff} \quad u'(y_t + m_{t-1})/u'(\bar{y} + \bar{m}) > \beta/\bar{p} > 1. \tag{A.5}$$

As a result, if y_t, m_{t-1}, or both drop significantly far away from their equilibrium values such that $u'(y_t + m_{t-1})/u'(\bar{y} + \bar{m})$ is larger than $\beta/\bar{p}$, it will be optimal to sell some milk in the local spot market, even when prices in the local spot market are a factor of β below prices offered by cooperatives. In that case, in the next period, income from the cooperative will be lower, and if the reduction in dairy income is large enough such that at time $t+1$,

$$\bar{p}u'(\bar{y} + \bar{m} - s_t) - \beta u'(\bar{y} + \bar{m}) > 0, \quad \textit{iff} \quad u'(\bar{y} + \bar{m} - s_t)/u'(\bar{y} + \bar{m}) > \beta/\bar{p} > 1, \tag{A.6}$$

---

[14] s_{t-1} might deviate from 0 if the increase in m_{t-1} is significantly large.

[15] s_{t-1} remains at 0 with a decrease in m_{t-1}.

farmers will continue selling some milk in the local market ($s_{t+1} > 0$). Note that because responses of s_t to changes in (y_t, m_{t-1}) in general are smaller than changes in (y_t, m_{t-1}) in themselves, condition (A.6) is much harder to satisfy than condition (A.5) unless there is a trend of change in milk production or net income.

Thus, in response to an unexpected drop in current nondairy income, y_t; predetermined milk production, m_{t-1}; or both, farmers who delivered all milk to the cooperative previously will *increase* the quantity of milk sold in the local market in the current period. This strategy, however, reduces future payments from the cooperative. Lower cash on hand in future periods may induce farmers to keep selling some of their milk in the local market. □

Farmers Sold All Milk in the Local Market, $\bar{s} = \bar{m}$

Now consider the case in which $\bar{p} > \beta$, so that the farmer sells all milk in the local market.

Lemma 3. *Consider the utility maximization problem given in* (1) *in the equilibrium when* $\bar{p} > \beta$ *such that* $\bar{s} = \bar{m}$*. Then, (1) if there is a decrease in current income,* $y_t < \bar{y}$*; a decrease in past milk production,* $m_{t-1} < \bar{m}$*; or both, then* $s_t = \bar{m}$*; (2) if there is an increase in* y_t*,* m_{t-1}*, or both, such that* $u'(y_t + \bar{p}\bar{m} + m_{t-1} - s_{t-1})/u'(\bar{y} + \bar{p}\bar{m}) < \beta/\bar{p}$*, then* $s_t < \bar{m}$*. Further, if the response* s_t *satisfies* $u'(\bar{y} + \bar{p}\bar{m} + \bar{m} - s_t)/u'(\bar{y} + \bar{m}) < \beta/\bar{p}$*, then also* $s_{t+1} < \bar{m}$*.*

Proof. By a similar analysis as Lemma 2 in the previous section, a decrease in y_t, m_{t-1}, or both will not deviate s_t from its equilibrium level, $\bar{m}$, because the positive first-order derivative (A.4) becomes more enforced. We consider the effect of an increase in current net income, $y_t > \bar{y}$; an increase in predetermined milk production, $m_{t+1} > \bar{m}$; or both. The first-order derivative, evaluated at $s_t = \bar{m}$ (so that the next period is not affected, leaving $s_{t+1} = \bar{m}$), is now[16]

$$\bar{p}u'(y_t + \bar{p}\bar{m} + m_{t-1} - s_{t-1}) - \beta u'(\bar{y} + \bar{p}\bar{m}) < 0 \quad \textit{iff} \quad u'(y_t + \bar{p}\bar{m} + m_{t-1} - s_{t-1})/u'(\bar{y} + \bar{p}\bar{m}) < \beta/\bar{p} < 1.$$

As a result, if the increase in y_t, m_t, or both is large enough that $u'(y_t + \bar{p}\bar{m} + m_{t-1} - s_{t-1})/u'(\bar{y} + \bar{p}\bar{m})$ is less than $\beta/\bar{p}$, it will be optimal to sell some milk to the cooperative instead of selling all of it in the local market. Thus, for farmers selling all milk in the local market, a sufficient increase in net income, milk production, or both will induce them to start selling milk to the cooperative.

Moreover, farmers will continue to prefer selling some milk to the cooperative in the next period, $t+1$, if the deferred payment for milk delivered at time t is above a certain threshold such that

$$\bar{p}u'(\bar{y} + \bar{p}\bar{m} + \bar{m} - s_t) - \beta u'(\bar{y} + \bar{m}) < 0, \quad \textit{iff} \quad u'(\bar{y} + \bar{p}\bar{m} + \bar{m} - s_t)/u'(\bar{y} + \bar{m}) < \beta/\bar{p} < 1.$$

In that case, an increase in either past milk production or current nondairy income continues to improve the quantity of milk delivered to the cooperative, including in future periods. □

[16] s_{t-1} increases with m_{t-1}, but not necessarily equals to m_{t-1}, since farmers might want to smooth consumption by delivering some milk to the cooperative.

Farmers Are Indifferent as to Where to Sell Their Milk, $\bar{s} \in (0, \bar{m})$

With $\bar{p} = \beta$, farmers are indifferent about where to sell milk, and $\bar{s}$ could be in anywhere between 0 and $\bar{m}$ but is unstable.

Lemma 4. *Consider the utility maximization problem given in* (1) *in the equilibrium when* $\bar{p} = \beta$ *such that* $\bar{s} \in (0, \bar{m})$. *Then, (1) if there is a decrease in current income,* $y_t < \bar{y}$*; a decrease in past milk production,* $m_{t-1} < \bar{m}$*; or both, then* $\bar{s} < s_t < s_{t+1} < \cdots$*, but* $\{s_t\}$ *does not necessarily converge to* $\bar{m}$*, depending on the property of the underlying utility function; (2) if there is an increase in* y_t*,* m_{t-1}*, or both, then* $\bar{s} > s_t > s_{t+1} > \cdots$*, and* $\{s_t\}$ *does not necessarily converge to 0.*

Proof. Whenever there is a shift in (y_t, m_{t-1}), s_t will respond in the opposite direction to maintain a 0 first-order derivative (A.1). This effect will pass along into the future and is decreasing return to scale, which can be seen by comparing the partial derivatives of the first-order condition (setting A.1 equal to 0) with respect to s_t and s_{t-1}, respectively. In this sense, the new equilibrium, $\bar{s}$, does not necessarily converge to 0 or $\bar{m}$ due to a one-time shift in (y_t, m_{t-1}), all else equal. □

APPENDIX B: SEMIPARAMETRIC ESTIMATOR AND ITS IMPLEMENTATION

Appendix B provides a brief introduction to the local linear kernel estimator and gives asymptotic properties of $\hat{\beta}$ and $\hat{M}(z)$, based on which we construct confidence intervals. Consistent estimators of the covariance matrices and bandwidth selection method are also provided.

The Local Linear Kernel Estimator

A local linear kernel estimator is used throughout this paper due to its advantages over the Nadaraya-Watson estimator (bias reduction, better behavior at boundary, ability to estimate derivatives, and so on, see Fan, 1992, 1993). To facilitate understanding of the estimation procedure, here we take Equation (5), the unfeasible local linear estimator, as an example. We denote the pseudo regressand as $y_{it}^* \equiv y_{it} - \alpha_i - x_{it}'\beta$ and put in vector form $Y^* = (y_{11}^*, \cdots, y_{1T}^*, y_{21}^*, \cdots, y_{nT}^*)'$. Given $\mathrm{E}(y_{it}^*|z_{it}) = m(z_{it})$, for z_{it} of size 2×1, the unfeasible local linear estimators for $M(z) \equiv \left(m(z), (H\dot{m}(z))'\right)'$, and $m(z)$ are given as

$$M_{\alpha,\beta}(z) = \arg\min_{\theta \in \mathbb{R}^3} \sum_{i=1}^{n} \sum_{t=1}^{T} \left(y_{it}^* - \theta_0 - (z_{it}' - z')(\theta_1, \theta_2)'\right)^2 K\left(H^{-1}(z_{it} - z)\right)$$
$$= \arg\min_{\theta \in \mathbb{R}^3} \left(Y^* - \vec{Z}(z)\theta\right)' \mathbf{K}_H(z) \left(Y^* - \vec{Z}(z)\theta\right), \tag{B.1}$$

where $\vec{Z}(z) \equiv \left(Z_{11}(z), \cdots, Z_{1T}(z), Z_{21}(z), \cdots, Z_{nT}(z)\right)'$, $Z_{it}(z) \equiv \left(1, H^{-1}(z_{it} - z)'\right)'$, and $\mathbf{K}_H(z) \equiv diag\left\{K_H(z_{it} - z)\right\}_{i,t=1}^{n,T}$. $K_H(z) \equiv |H|^{-1}K(H^{-1}z)$, where $K(z)$ is a kernel function on $\mathbb{R}^2$, $H = diag(h_1, h_2)$ is matrix of bandwidths, and $|H|$ is the determinant of H.

Minimization of (B.1) gives $M_{\alpha,\beta}(z) = \left(\vec{Z}(z)'\mathbf{K}_H(z)\vec{Z}(z)\right)^{-1}\vec{Z}(z)'\mathbf{K}_H(z)Y^* \equiv S(z)Y^*$ and $m_{\alpha,\beta}(z) = s(z)'Y^*$, where $s(z)' \equiv e'S(z)$ and $e = (1,0,0)'$. Here, $S(z)$ and $s(z)'$ are the local linear estimator premultipliers in Equations (5) and (7).

Asymptotic Properties of $\hat{\beta}$ and $\hat{M}(z)$

Asymptotic properties of $\hat{\beta}$ and $\hat{M}(z)$ are established by Su and Ullah (2006) in Theorems 3.1 and 3.2 respectively, under assumptions A1–A7. For $\hat{\beta}$, let $\tilde{x}_{it} = x_{it} - \mathrm{E}(x_{it}|z_{it})$, $\vec{x}_{it} = x_{it} - \left(s(z_{it})'X\right)'$, and $\hat{v}_{it} = y_{it} - x_{it}'\hat{\beta} - \hat{m}(z_{it}) - \hat{\alpha}_i$, so that we have

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} \mathcal{N}\left(0, \Sigma\right), \tag{B.2}$$

where $\Sigma = \Phi^{-1}\Omega\Phi^{-1}$, $\Phi = \sum_t \mathrm{E}\left(\tilde{x}_{it}(\tilde{x}_{it} - \sum_s \tilde{x}_{is}/T)'\right)$, and $\Omega = \sum_s \sum_t \mathrm{E}\left(\tilde{x}_{it}(\tilde{x}_{is} - \sum_l \tilde{x}_{il}/T)'v_{it}v_{is}\right)$. A consistent estimator for Σ is given by $\hat{\Sigma} = \hat{\Phi}^{-1}\hat{\Omega}\hat{\Phi}^{-1}$, where $\hat{\Phi} = n^{-1}\sum_i \sum_t \vec{x}_{it}(\vec{x}_{it} - \sum_l \vec{x}_{il}/T)'$, and $\hat{\Omega} = n^{-1}\sum_i \sum_t \sum_s \vec{x}_{it}(\vec{x}_{is} - \sum_l \vec{x}_{il}/T)'\hat{v}_{it}\hat{v}_{is}$.

For $\hat{M}(z)$, let $\bar{f}(z)=\sum_t^T f_t(z)$, $\tilde{v}_{it}=v_{it}-T^{-1}\sum_{s=1}^T v_{is}$, $\sigma_t^2(z)=\mathrm{E}(\tilde{v}_{it}^2|z_{it}=z)$, and $\bar{\sigma}^2(z)=\sum_{t=1}^T \sigma_t^2(z)f_t(z)$, so that we have

$$\sqrt{n|H|}\left(\hat{M}(z)-M(z)-Q^{-1}\begin{pmatrix}\frac{\bar{f}(z)}{2}tr(\int_{\mathbb{R}^2}uu'K(u)\mathrm{d}uH\ddot{m}(z)H)\\ 0\end{pmatrix}\right)\xrightarrow{d}\mathcal{N}\left(0,Q^{-1}\Gamma Q^{-1}\right), \tag{B.3}$$

where $\ddot{m}(z)$ is the second order derivative matrix of $m(\cdot)$ at z,

$$Q=\bar{f}(z)\begin{pmatrix}1 & 0'\\ 0 & \int_{\mathbb{R}^2}uu'K(u)\mathrm{d}u\end{pmatrix}\quad\text{and}\quad\Gamma=\bar{\sigma}^2(z)\begin{pmatrix}\int_{\mathbb{R}^2}K(u)^2\mathrm{d}u & 0'\\ 0 & \int_{\mathbb{R}^2}uu'K(u)\mathrm{d}u\end{pmatrix}.$$

The asymptotic normal distribution derived in Equation (B.3) can be used to obtain pointwise confidence intervals for the nonparametric estimator $\hat{M}(\cdot)$. As argued by Härdle and Linton (1994), in practice, it is usual to ignore the bias term since it usually depends on higher-order derivatives of the regression function, in our case, $\ddot{m}(\cdot)$. Thus we choose a smaller bandwidth than that used in the cross-validation method to make the bias relatively small. In our application, we adopt the commonly used Epanechnikov product kernel, $K(u)\equiv\prod_{i=1}^2 k(u_i)$, where $k(u)=0.75(1-u^2)1(|u|\le 1)$. Let $\mu_i\equiv\int_{\mathbb{R}}u^ik(u)\mathrm{d}u$ and $\nu_i\equiv\int_{\mathbb{R}}u^ik(u)^2\mathrm{d}u$. Then, $\int_{\mathbb{R}^2}uu'K(u)\mathrm{d}u=\begin{pmatrix}\mu_2\mu_0 & \mu_1^2\\ \mu_1^2 & \mu_2\mu_0\end{pmatrix}=\begin{pmatrix}1/5 & 0\\ 0 & 1/5\end{pmatrix}$, and $\int_{\mathbb{R}^2}K(u)^2\mathrm{d}u=\nu_0^2=3/5$.

By Equation (B.3), it is easy to obtain the elementwise asymptotic normal distribution for $\hat{m}(\cdot)$ and the slope estimator, $\hat{\dot{m}}_i(\cdot)$, for $i=1,2$,

$$\sqrt{n|H|}\left(\hat{m}(z)-m(z)-\frac{\mu_2}{2}\sum_{i=1}^2 h_i^2\ddot{m}_{ii}(z)\right)\xrightarrow{d}\mathcal{N}\left(0,\frac{\nu_0^2\bar{\sigma}^2(z)}{\bar{f}^2(z)}\right)\quad\text{and} \tag{B.4}$$

$$\sqrt{nh_i^2|H|}\left(\hat{\dot{m}}_i(z)-\dot{m}_i(z)\right)\xrightarrow{d}\mathcal{N}\left(0,\frac{\nu_0\nu_2\bar{\sigma}^2(z)}{\mu_2^2\bar{f}^2(z)}\right), \tag{B.5}$$

where $\bar{f}(z)=\sum_t^T f_t(z)$, $\bar{\sigma}^2(z)=\sum_{t=1}^T\sigma_t^2(z)f_t(z)$, $\dot{m}_i(z)$ is the ith element in the first-derivative vector of $m(\cdot)$ at z, and $\ddot{m}_{ii}(z)$ is the ith diagonal element in the second-derivative matrix of $m(\cdot)$ at z.

Due to limited observations, it is hard to estimate the variance function for each period. For simplicity, we assume homoscedasticity and the same joint density function of the covariates across time, that is, $\sigma_t^2(z)=\sigma_t^2<\infty$ and $f_t(z)=f(z)$. Then we have $\frac{\bar{\sigma}^2(z)}{\bar{f}^2(z)}=\frac{\sum_{t=1}^T\sigma_t^2}{T^2f(z)}$. A consistent estimator for σ_t^2 is $\hat{\sigma}_t^2=n^{-1}\sum_{i=1}^n\hat{\tilde{v}}_{it}^2$, where $\hat{\tilde{v}}_{it}=\hat{v}_{it}-T^{-1}\sum_{s=1}^T\hat{v}_{is}$. We use the consistent Parzen-Rosenblatt density estimator, $\hat{f}(z)$, to approximate $f(z)$, where $\hat{f}(z)\equiv(nT|H|)^{-1}\sum_{i=1}^n\sum_{t=1}^T K\left(H^{-1}(z_{it}-z)\right)$. $K(\cdot)$ is the Epanechnikov product kernel and H is obtained using the cross-validation method. Thus, based on (B.4), (B.5), and estimators for the covariances, the pointwise 95 percent confidence intervals of $\hat{m}(z)$ and $\hat{\dot{m}}_i(z)$ are constructed, respectively, as $\mathrm{CI}(z)=[\hat{m}(z)-2s(z),\hat{m}(z)+2s(z)]$ and $\mathrm{CI}_i(z)=[\hat{\dot{m}}_i(z)-2s_i(z),\hat{\dot{m}}_i(z)+2s_i(z)]$, where $s(z)=\left((n\hat{h}_1\hat{h}_2T^2\hat{f}(z))^{-1}\nu_0^2\sum_{t=1}^T\hat{\sigma}_t^2\right)^{1/2}$ and $s_i(z)=\left((n\hat{h}_1\hat{h}_2\hat{h}_i^2\mu_2^2T^2\hat{f}(z))^{-1}\nu_0\nu_2\sum_{t=1}^T\hat{\sigma}_t^2\right)^{1/2}$ for $i=1,2$.

Bandwidth Selection

In nonparametric estimation, bandwidth selection is crucial in order to obtain a good estimate. We choose bandwidths using the leave-one-out cross-validation method. Specifically, given that $z = (z_1, z_2)'$ consists of two variables, let $h \equiv (h_1, h_2) = (c_1\hat{s}_{z_1}(nT)^{-1/6}, c_2\hat{s}_{z_2}(nT)^{-1/6})$, where $\hat{s}_{z_i}$ is the sample deviation of z_i for $i = 1, 2$. We choose h to be

$$\tilde{h} = \arg\min_{h\in\mathbf{H}} CV(h) \equiv \arg\min_{h\in\mathbf{H}} \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y_{it} - \hat{m}^{-it}(z_{it})\right)^2, \tag{B.6}$$

where $\hat{m}^{-it}(z_{it})$ is the local linear estimator for $\mathrm{E}(y_{it}|z_{it} = z)$ obtained by deleting the observation z_{it}. Consistent with Su and Ullah (2006), $\tilde{h}$ is derived by a grid search over the mesh grid $\mathbf{H}$ of two dimensions, with each constructed by an interval $[0.1\hat{s}_{z_i}(nT)^{-1/6}, 10\hat{s}_{z_i}(nT)^{-1/6}]$ over 20 steps for $i = 1, 2$. Thus, $\mathbf{H}$ consists of 400 points in total. We denote the constants associated with $\tilde{h}$ as $(\hat{c}_1, \hat{c}_2)$. Note that there is an asymptotic bias term in $\hat{M}(z)$, as is shown in (B.3). It is more convenient to undersmooth a little bit (that is, let $nh^6 \to 0$ as $n \to \infty$) than to estimate the unknown term $\ddot{m}(z)$ in the bias. Thus, in practice, we choose bandwidths $\hat{h} \equiv (\hat{c}_1\hat{s}_{z_1}(nT)^{-1/6-0.01}, \hat{c}_2\hat{s}_{z_2}(nT)^{-1/6-0.01})$.

To implement Parzen-Rosenblatt density estimators, we choose a cross-validation method that minimizes the estimate of the integrated squared error, $\int(\hat{f}(z) - f(z))^2 \mathrm{d}z$. In particular,

$$\begin{aligned}\hat{h}_d = \arg\min_{h} CV_d(h) \equiv \arg\min_{h} \Bigg\{ &\frac{1}{(nT)^2|H|}\sum_{i_1=1}^{n}\sum_{t_1=1}^{T}\sum_{i_2=1}^{n}\sum_{t_2=1}^{T}\bar{K}\left(H^{-1}(z_{i_1t_1} - z_{i_2t_2})\right) \\ &- \frac{2}{nT(nT-1)|H|}\sum_{i_1=1}^{n}\sum_{t_1=1}^{T}\sum_{(i_2,t_2)\neq(i_1,t_1)} K\left(H^{-1}(z_{i_1t_1} - z_{i_2t_2})\right)\Bigg\},\end{aligned} \tag{B.7}$$

where $\bar{K}(\cdot)$ is the convolution function of the kernel $K(\cdot)$. For the Epanechnikov product kernel, $\bar{K}(u) = \bar{k}(u_1)\bar{k}(u_2)$, where $\bar{k}(u) = \int k(x)k(x-u)\mathrm{d}x = \frac{3}{160}(2-|u|)^3(u^2 + 6|u| + 4)1(|u| \leq 2)$. $\hat{h}_d$ is obtained using *fminsearch* in MATLAB R2016a with initial value $h_0 = (\hat{s}_{z_1}, \hat{s}_{z_2})(nT)^{-1/6}$. This command uses the Nelder-Mead simplex direct search algorithm developed by Lagarias et al. (1998).

REFERENCES

Casaburi, L., and R. Macchiavello. 2015. *Firm and Market Response to Saving Constraints: Evidence from the Kenyan Dairy Industry.* CEPR Discussion Paper No. DP10952.Washington, DC: Center for Economic and Policy Research.

Collins, D., J. Morduch, S. Rutherford, and O. Ruthven. 2009. *Portfolios of the Poor: How the World's Poor Live on $2 a Day.* Princeton, NJ, US: Princeton University Press.

Dean, M., and A. Sautmann. 2016. *Credit Constraints and the Measurement of Time Preferences.* Department of Economics Working Paper. Providence, RI, US: Brown University.

Epper, T., 2015. *Income Expectations, Limited Liquidity, and Anomalies in Intertemporal Choice.* Economics Working Paper. St. Gallen, Switzerland: Univeristy of St. Gallen.

Fan, J., 1992. "Design-adaptive Nonparametric Regression." *Journal of the American Statistical Association* 87: 998–1004.

Fan, J., 1993. "Local Linear Regression Smoothers and their Minimax Efficiencies." *Annals of Probability* 21: 196–216.

Gulati, A., N. Minot, C. Delgado, and S. Bora. 2007. "Growth in High-value Agriculture in Asia and the Emergence of Vertical Links with Farmers," In *Global Supply Chains, Standards and the Poor: How the Globalization of Food Systems and Standards affects Rural Development and Poverty*, edited by J. F. Swinnen 91–108. Cambridge, MA, US: CAB International.

Härdle, W., and O. Linton. 1994. "Applied Nonparametric Methods." *Handbook of Econometrics* 1069: 2295–2339.

Ide, V. N., W. Janssens, B. Kramer, and M. Van der List. 2016. "Coping with Health Shocks: Remittances, Insurance or Both? Evidence from Financial and Health Diaries Data in Kenya." Paper presented at the 2016 NOVAFRICA Conference on Economic Development in Africa, Lisbon, Portugal, July 14–15.

Jack, W., M. Kremer, J. de Laat, and T. Suri. 2016. *Borrowing Requirements, Credit Access, and Adverse Selection: Evidence from Kenya.* NBER Working Paper No. 22686. Cambridge, MA, US: National Bureau of Economic Research.

Janssens, W., B. Kramer, M. Van der List, and D. Pap. 2013. *Financial and Health Diaries: A Yearlong Weekly Panel of Farming Households in Nigeria and Kenya.* Amsterdam: Amsterdam Institute for International Development and PharmAccess Foundation.

Janssens, W., B. Kramer, and L. Swart. 2016. *Be Patient when Measuring Hyperbolic Discounting: Stationarity, Time Consistency and Time Invariance in a Field Experiment.* Tinbergen Institute Discussion Paper. Amsterdam: Tinbergen Institute.

Kramer, B., and D. Kunst. 2016. "Intertemporal Allocations of Gifts Versus Earnings: Evidence from a Field Experiment in Kenya." Unpublished, International Food Policy Research Institute, Washington, DC.

Lagarias, J., J. Reeds, H. Wright, and P. Wright. 1998. "Convergence Properties of the Neldermead Simplex Method in Low Dimensions." *Journal of Optimization* 9 (1): 112–117.

Minot, N., and B. Sawyer. 2014. "Contract Farming in Developing Countries: Review of the Evidence." Paper prepared for the Investment Climate Unit of the International Finance Corporation. Washington, DC.

Reardon, T., C. B. Barrett, J. A. Berdegue´, and J. F. Swinnen. 2009. "Agrifood Industry Transformation and Small Farmers in Developing Countries." *World Development* 37 (11): 1717–1727.

Su, L., and A. Ullah. 2006. "Profile Likelihood Estimation of Partially Linear Panel Data Models with Fixed Effects. *Economics Letters* 92 (1): 75–81.

Zimmerman, F. J., and M. R. Carter. 2003. "Asset Smoothing, Consumption Smoothing and the Reproduction of Inequality under Risk and Subsistence Constraints. *Journal of Development Economics* 71 (2): 233–260.

RECENT IFPRI DISCUSSION PAPERS

For earlier discussion papers, please go to www.ifpri.org/publications/discussion_papers.
All discussion papers can be downloaded free of charge.

1601. *A chicken and maize situation: The poultry feed sctor in Ghana.* Kwaw S. Andam, Michael E. Johnson, Catherine Ragasa, Doreen S. Kufoalor, and Sunipa Das Gupta, 2017.

1600. *Gender justice and food security in India: A review.* Nitya Rao, Mamata Pradhan, and Devesh Roy, 2017.

1599. *Cities and rural transformation: A spatial analysis of rural youth livelihoods in Ghana.* Xinshen Diao, Peixun Fang, Eduardo Magalhaes, Stefan Pahl, and Jed Silver, 2017

1598. *The changing structure of Africa's economies.* Xinshen Diao, Kenneth Harttgen, and Margaret McMillan, 2017.

1597. *Perspectives on the role of the state in economic development: Taking stock of the "Developmental State" after 35 years.* Jordan Kyle, 2017.

1596. *Imputing nutrient intake from foods prepared and consumed away from home and other composite foods: Exploring extensions of the Subramanian–Deaton cost per calorie approach.* Dena M. Mwangi, John L. Fiedler, and Celeste Sununtnasuk, 2017.

1595. *Estimating spatial basis risk in rainfall index insurance: Methodology and application to excess rainfall insurance in Uruguay*. Francisco Ceballos, 2016.

1594. *The effect of land inheritance on youth employment and migration decisions: Evidence from rural Ethiopia.* Katrina Kosec, Hosaena Ghebru, Brian Holtemeyer, Valerie Mueller, and Emily Schmidt, 2016.

1593. *How headquarters relocation is affected by rising wages and ownership: Evidence from China's Annual Survey of Industrial Enterprises, 1999–2008.* Qingtao Wang, Kevin Chen, Longwen Chiang, and Xuanli Xie, 2016.

1592. *External shocks, food security, and development: Exploring scenarios for Central America.* Eugenio Diaz-Bonilla, Valeria Piñeiro, and Pablo Elverdin, 2016.

1591. *The role of learning in technology adoption: Evidence on hybrid rice adoption in Bihar, India.* Jared Gars and Patrick S. Ward, 2016.

1590. *Building effective clusters and industrial parks.* Xiaobo Zhang, 2016.

1589. *Nepal's 2072 federal constitution: Implications for the governance of the agricultural sector.* Jordan Kyle and Danielle Resnick, 2016.

1588. *Entitlement fetching or snatching?: Effects of arbitrage on India's public distribution system.* Suman Chakrabarti, Avinash Kishore, and Devesh Roy, 2016.

1587. *Changing preferences through experimental games: Evidence from sanitation and hygiene in Tamil Nadu.* Yaniv Stopnitzky, 2016.

1586. *Climate change, agriculture, and adaptation in the Republic of Korea to 2050: An integrated assessment.* Nicola Cenacchi, Youngah Lim, Timothy Sulser, Shahnila Islam, Daniel Mason-D'Croz, Richard Robertson, Chang-Gil Kim, and Keith Wiebe, 2016.

1585. *The dark side of competition: Gender differences.* Simon Chang, Kamhon Kan, and Xiaobo Zhang, 2016.

1584. *Harnessing net primary productivity data for monitoring sustainable development of agriculture.* Nathaniel P. Robinson, Cindy M. Cox, and Jawoo Koo, 2016.

1583. *Effects of agricultural mechanization on smallholders and their self-selection into farming: An insight from the Nepal Terai.* Hiroyuki Takeshima, Rudra Bahadur Shrestha, Basu Dev Kaphle, Madhab Karkee, Suroj Pokhrel, and Anjani Kumar, 2016.

1582. *The agricultural sector as an alternative to illegal mining in Peru: A case study of Madre de Dios.* Valeria Piñeiro, Jenny Thomas, and Pablo Elverdin, 2016.

1581. *What drives diversification of national food supplies?: A cross-country analysis.* Samira Choudhury and Derek Headey, 2016.

1580. *Identity, household work, and subjective well-being among rural women in Bangladesh.* Greg Seymour and Maria S. Floro, 2016.

INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE

www.ifpri.org

IFPRI HEADQUARTERS

2033 K Street, NW
Washington, DC 20006-1002 USA
Tel.: +1-202-862-5600
Fax: +1-202-467-4439
Email: ifpri@cgiar.org