Liquid Milk: Cash Constraints and Recurring Savings among Dairy Farmers in Kenya

Xin Geng, Wendy Janssens and Berber Kramer*

*Markets, Trade and Institutions Division, International Food Policy Research Institute, b.kramer@cgiar.org


Copyright 2018 by Xin Geng, Wendy Janssens and Berber Kramer. All rights reserved. Readers may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies.
Liquid Milk: Cash Constraints and Recurring Savings among Dairy Farmers in Kenya

Xin Geng∗, Wendy Janssens†, and Berber Kramer‡

January, 2018

Abstract

We analyze whether cash constraints affect recurring savings, using high-frequency panel data for Kenyan dairy farmers. These farmers sell milk to a cooperative, which pays once a month, and to market vendors, who pay a lower price but upon delivery. Using semiparametric estimation methods, we find that farmers save a smaller share of their dairy income and sell more to market vendors in periods with less cash on hand and in weeks with uninsured health shocks. We conclude that cash constraints induce side-selling of milk, which can lower output prices that farmers receive and destabilize recurring savings with the cooperative.

JEL codes: D14; D15; O16; Q13.

Keywords: liquidity constraints; savings; panel data; semiparametric estimation.

∗Markets, Trade and Institutions Division, International Food Policy Research Institute (IFPRI), Washington, DC.
†Department of Economics, VU University Amsterdam; Tinbergen Institute.
‡Corresponding author, b.kramer@cgiar.org. Markets, Trade and Institutions Division, International Food Policy Research Institute (IFPRI), Washington, DC. We are grateful to the PharmAccess Foundation and Amsterdam Institute for International Development for funding the Health and Financial Diaries project; to the Synovate Kenya team for their dedication to data collection; and to our respondents for their time and patience. We are thankful for suggestions made by Mark Dean, Glenn Harrison, Anja Sautmann, and Charles Sprenger, and for feedback from participants in seminars at IFPRI, DIW Berlin, Tinbergen Institute, the 2017 ASSA meetings in Chicago and the 8th Development Economics Workshop at Wageningen University. This work was undertaken as part of the CGIAR Research Programs on Policies, Institutions and Markets (PIM), and Agriculture for Nutrition and Health (A4NH). The opinions expressed here belong to the authors, and do not necessarily reflect those of A4NH, PIM, IFPRI, or CGIAR.
1 Introduction

The rural poor rely for a large share of their incomes on spot markets, generating frequent inflows of cash, for instance from casual labor, small businesses, and sales of agricultural produce such as milk or vegetables. Accumulating such income is important to save for lump-sum expenses and future investments, but is challenging since access to savings accounts and other financial services is limited particularly in rural areas (Demirgüç-Kunt et al., 2015). To meet their savings goals, the rural poor accumulate part of their income through recurring deposits in informal commitment savings devices. For instance, they save through money collectors (Collins et al., 2009), rotating savings and credit associations (ROSCAs, Gugerty, 2007), or, when selling their produce, they can defer payments by selling to cooperatives, which typically pay in the future and in bulk instead of upon delivery (Casaburi and Macchiavello, 2016).

Participating in these informal commitment savings devices on a recurring and consistent basis can help accumulate wealth for at least two reasons. First, building strong savings habits is important in overcoming temptations to spend on current needs and pressures to share (Karlan et al., 2014). Second, the sustainability of these savings devices falls or stands by the reliability of members’ deposits; ROSCAs will collapse if too many members default in a given week, and a cooperative will not be able to negotiate good prices when the supply of output is unpredictable (Van den Brink and Chavas, 1997; Barham and Chitemi, 2009). At the same time, in the presence of income risk and liquidity constraints, savings are thought to respond to cash on hand as an insurance device (Deaton, 1991), compromising wealth accumulation for future investments.1

We therefore study to what extent variation over time in recurring deposits is caused by varia-

---

1 Credit constraints influence behavior not only in developing but also in developed countries. For instance, recent evidence from the US shows that liquidity constrained subprime borrowers consumers tend to have higher rates of mortgage default (Anderson and Dokko, 2016), and that lifting banking restrictions—relaxing credit constraints—raises college enrollment (Sun and Yannelis, 2016).
tion in cash on hand. Several studies find that variation in financial resources influences decision-making in experiments that elicit preferences over smaller-sooner versus later-larger payoffs. For instance, Carvalho et al. (2016) find an increase in present bias just prior to payday among low-income households from the United States, and Janssens et al. (2017) find that households make relatively less patient decisions when they have fewer financial resources especially when facing borrowing constraints. Nonetheless, recurring deposits in informal commitment savings devices may be less sensitive to variation in cash on hand if they become automated habits, which are less susceptible to change than non-habitual behaviors (Ouellette and Wood, 1998). Moreover, social norms and bylaws require members to contribute or deliver produce—in other words, to deposit—on a regular, recurring basis (Casaburi and Macchiavello, 2015).

To analyze whether fluctuating levels of cash on hand influence recurring deposits, we use panel data for members of a dairy cooperative in Kenya. As many other cooperatives, this institution automates the savings decision for farmers who sell their milk to the cooperative: by paying once a month, it defers a farmer’s milk payment with an average of 25 days. As a return on this delay in payment, the cooperative tends to offer higher milk prices than the spot market; the monthly cooperative price in our study area is on average an estimated 19 to 35 percent higher than milk prices in the local market. We hence model the decision to sell milk to the cooperative as an informal savings device that yields a higher return compared with buffer stock assets such as cash, jewelry or small livestock. Within the context of this model, we predict how the share of sales to the cooperative—which we interpret as a recurring deposit—depends on cash on hand.

We analyze the relation between these recurring deposits and cash on hand 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. To proxy cash on hand, we measure cash flows from non-dairy activities in the last week, and milk production in the last

---

2 The cooperative pays around the 11th of the next month for all milk delivered in the previous month.
month (as an exogenous determinant for the payment of milk sold to the cooperative in the previous month). Our model predicts that these two variables have nonlinear and nonadditive effects. We therefore estimate the effects of these variables using semiparametric estimation methods, allowing cash on hand to enter the model nonparametrically. As an alternative proxy for cash needs, we also study the effect of health shocks for households without and with insurance coverage.

Controlling for various covariates as well as household and center-month fixed effects, our semiparametric estimates indicate that households reduce the share of dairy sales to the cooperative, which defers payments, in weeks that follow periods of low levels of milk production and non-dairy income. In other words, when households have less cash on hand from either dairy farming or non-dairy activities, they earn a larger share of dairy sales in cash from local market vendors. Moreover, households sell a relatively larger share of milk in the market in weeks that a family member is ill or injured. We do not observe this effect in weeks that medical bills are covered by health insurance. These findings suggest that households sell in the market in order to satisfy urgent cash needs. We find gender heterogeneity in how this affects the average price at which farmers sell milk.

This paper contributes to the existing literature in three ways. First, using high-detail high-frequency panel data, we analyze recurring savings deposits outside the stylized context of a laboratory. We are able to observe variation over time in where households sell their milk—determining whether they save their dairy income—and relate this to variation in cash on hand, which allows us to test whether financial resources affect savings decisions. The finding that cash on hand affects savings rates is consistent with recent field experiments showing that individuals with less wealth reveal less patient preferences (Dean and Sautmann, 2016; Janssens et al., 2017), but we show this for recurring savings as opposed to allocations of an unexpected experimental windfall. For two different cooperatives, Casaburi and Macchiavello (2016) and Kramer and Kunst (2017) elicit preferences for when to receive milk payments, but in a one-time experimental setting. They also do not measure cash on hand through panel data and hence do not shed light on how financial resources affect recurring deposits.
Second, we test these predictions using a semiparametric model (Su and Ullah, 2006) that has to our best knowledge not been applied previously on microeconomic panel data. Our conceptual framework predicts the existence of threshold effects and nonlinearities in the relation between cash on hand and recurring savings, and two indicators of cash on hand—income from dairy farming and income from other activities—could be non-fungible. The model allows for both variables to enter nonparametrically, without presuming any functional form for these variables, so that potential misspecification cannot bias the results. We find that the semiparametric estimates provide a more nuanced and often very different picture than a fully parametric model with a linear interaction term for lagged milk production and lagged non-dairy income, underscoring the added value of using of semiparametric models when two variables have potential nonlinear interaction effects.

Third, we provide a potential explanation for why farmers engage in side-selling. Cooperatives and other aggregators can generate significant welfare gains by aggregating numerous smallholder farmers into collective groups, allowing them to access output markets for high-value agricultural commodities such as fruits, vegetables and milk (Gulati et al., 2007; Reardon et al., 2009). However, cooperatives can only leverage economies of scale when members deliver reliable volumes of output, without side-selling their produce in the local market (Bhuyan, 2007; Wollni and Fischer, 2014), and despite bylaws requiring members to deliver their output to the cooperative, side-selling occurs frequently. Although typically attributed to differences between cooperative and market prices (Minot and Sawyer, 2014), our findings suggest cash constraints play an important role as well. Variation in cash on hand destabilizes farmers’ willingness to defer milk payments, affecting the volumes of milk supplied to the cooperative.

The remainder of this paper is structured as follows. The next section presents a conceptual framework that we will use to derive our main hypotheses. Section 3 describes the data and semiparametric estimation method that we will use to test these hypotheses. Section 4 presents semiparametric results and compares our results with 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 and discusses areas for future research.

### 2 Conceptual framework

This section describes the choice to save by selling milk to the cooperative from a conceptual point of view, building on Deaton (1991), who models consumption-savings decisions in the presence of risk and liquidity constraints. Deaton shows that liquidity-constrained consumption paths violate the permanent income hypothesis in that consumption does not only depend on permanent income, but also on cash on hand at the time of the consumption-savings decision. Moreover, although the model predicts hand-to-mouth consumption at low levels of cash on hand, it also predicts savings through low-return assets at higher levels of cash on hand, for which consumers with access to credit would choose not to save. Liquidity-constrained consumers accumulate such buffer stock assets to self-insure future consumption against bad income draws.

Our conceptual framework extends this model by introducing the choice between selling to a cooperative, which defers payments until the next period, versus local market vendors who pay a lower price but upon delivery. This savings device generates a higher return than the buffer stock through which consumers accumulate precautionary savings in the absence of the cooperative, and it follows intuitively that consumers will save first by selling to the cooperative, while the buffer stock fulfills any remaining savings demand. We will show that savings with the cooperative respond to cash on hand in a similar way as Deaton’s buffer stock savings: when cash on hand is high, farmers sell more to the cooperative in order to save, whereas lower levels of cash on hand induce them to sell in the market in order to obtain immediate cash. The goal of our empirical analyses is to test that prediction.

Formally, consider the following infinite-horizon framework with periods \( t \in \{0, \ldots, \infty\} \). As in the Deaton model, every period \( t \), farmers earn exogenous income (from non-dairy activities) \( y_t \),
and they can save an amount equal to $s^a_t \geq 0$ through a buffer stock with net returns equal to $r^a$. In addition, the farmer produces milk with market value $m_t$. When selling the milk to vendors in the local market, the farmer receives this amount immediately, in period $t$. Farmers can however decide to save some or all dairy income by selling to the cooperative, which defers payments to the next period $t+1$, but pays a higher price. Define the amount of dairy income saved by selling to the cooperative as $s^c_t \in [0, m_t]$, and the net return on these savings as $r^c > 0$.\(^3\)

We can define the amount of cash on hand, $A_t$, available for period-$t$ consumption, $c_t$, as

$$A_t = y_t + (1 + r^a)s^a_{t-1} + m_t + (1 + r^c)s^c_{t-1}.$$  \(1\)

Cash on hand includes income from non-dairy activities, $y_t$, any savings in the liquid asset from the previous period, $(1 + r^a)s^a_{t-1}$, the market value of period-$t$ milk production, $m_t$, and payments for milk delivered to the cooperative in the previous period, $(1 + r^c)s^c_{t-1}$. We assume that non-dairy income, $y_t$, and the market value of dairy production, $m_t$, are subject to risk; specifically, we assume that these are two stationary random variables with compact support. In period $t$, farmers have perfect information with regards to income and production draws for that period, and with regards to the distribution of future income and production draws.\(^4\)

Every period $t$, the farmer decides how much to save for period $t+1$ with the cooperative and in the liquid asset. Specifically, the farmer faces the following optimization problem:

$$\max_{s^a_t \geq 0, \ s^c_t \in [0, m_t]} \mathbb{E}_t \left\{ \sum_{\tau = t}^{\infty} (1 + \delta)^{t-\tau} v (A_t - s^a_\tau - s^c_\tau) \right\},$$  \(2\)

in which $A_t - s^a_t - s^c_t$ represents consumption $c_t$, as defined by a regular budget constraint, and in

\(^3\) This net return is the mark-up in price that the cooperative offers compared to market vendors. Note further that we assume the decision to consume milk at home, or any milk spillage, to be exogenous, and that we do not model that here for ease of presentation.

\(^4\) Empirically, non-dairy income and dairy production are likely to be serially correlated. This does not affect the optimal consumption pattern for liquidity-constrained consumers (Deaton, 1991), although in the limit, when income is a random walk, it is optimal for liquidity-constrained consumers to consume all their incomes and not to save.
which we define $\delta > 0$ as the discount rate and $v(\cdot)$ as the instantaneous utility function, assumed to be increasing, differentiable and strictly concave, meaning that farmers will want to smooth consumption.\(^5\)

Both savings devices require non-negative savings levels, $s^a_t \geq 0$ and $s^c_t \geq 0$. In other words, farmers cannot borrow against future income to shield consumption from negative income shocks.\(^6\)

Because of this cash constraint, farmers need to self-insure consumption by accumulating wealth. They can only do so using the buffer stock asset since they cannot save more than the market value of milk production with the cooperative, $s^c_t \leq m_t$.\(^7\) We also impose the condition that net returns on buffer stock savings are smaller than the discount rate, $r^a < \delta$. This seems to be a natural assumption for liquid or semi-liquid savings that come in the form of for instance cash, small livestock or jewelry.\(^8\) These returns are also smaller than returns on deferred payments from the cooperative, $r^a < r^c$, to reflect the empirical observation that in our study areas, cooperative prices are on average an estimated 19 to 35 percent higher than milk prices in the local market.\(^9\)

Appendix A describes the solution to this optimization problem. We find that the farmer will always prefer saving with the cooperative because the return on the buffer stock asset is smaller than the return on deferred milk payments from the cooperative, $r^a < r^c$. However, since farmers cannot save more than the market value of dairy production with the cooperative, $s^c_t \leq m_t$, savings

\(^5\) Deaton (1991) also assumes that the first derivative of utility is strictly convex to predict an increase in precautionary savings following a mean-preserving spread of income $y_t$. We do not require this condition as we study the effects of within-household variation in cash on hand instead of between-household variation in the degree of risk.

\(^6\) In practice, farmers might be able to borrow from their social network when in financial need, or request (at a fee) a cash advance from the cooperative. However, Geng et al. (2017) show that our target population is unable to fully smooth consumption in the presence of health shocks, and the diaries report only 11 transactions that involve a cash advance from the cooperative. This suggests that such informal insurance mechanisms are incomplete.

\(^7\) This means that they cannot purchase milk from others and deliver it under their own name. We impose this condition as we do not observe such milk trading patterns in our data.

\(^8\) Deaton (1991) points out that if the return on the liquid asset were equal to or greater than the discount rate, $r^a \geq \delta$, consumers could choose to accumulate assets indefinitely and use capital income to finance consumption. This would, however, make income shocks and hence borrowing constraints irrelevant.

\(^9\) We acknowledge that in practice, the net return on savings with the cooperative, $r^c$, might fluctuate across time, and that it might be lower than the net interest rate on assets in certain scenarios. On average, however, the assumption holds, and if the theory did not impose this assumption, farmers would always sell in the local market and save through assets for a better rate.
with the cooperative will be capped to this level at sufficiently high levels of cash on hand, and any remaining savings demand will be met with buffer stock savings. The threshold level of cash on hand at which this occurs depends on $m_t$. This yields the following prediction for the amount of dairy income that is saved by selling to the cooperative:

**Lemma 1.** Consider the utility maximization problem given in (2) subject to (1). Let $c_t = c(A_t, m_t)$ be a stationary stochastic optimum. There exist unique thresholds $A_1^*$ and $A_2^*(m)$ with $A_2^*(m) > A_1^* > 0$ such that

$$
c = c(A, m) = A \quad \text{and} \quad s^c = 0, \quad \text{if} \ A \leq A_1^*;
$$

$$
c = c(A, m) = A - s^c \quad \text{and} \quad 0 < s^c < m, \quad \text{if} \ A_1^* < A < A_2^*(m);
$$

$$
c = c(A, m) \leq A - m \quad \text{and} \quad s^c = m, \quad \text{if} \ A \geq A_2^*(m).
$$

Lemma 1 indicates that in theory, the share of dairy sales to the cooperative is related to milk production and non-dairy income in a nonlinear way for three reasons. First, changes in cash on hand need to pass a threshold $A_1^*$ for savings with the cooperative to start increasing in cash on hand. Second, once cash on hand exceeds a threshold $A_2^*$, which depends on current milk production, savings with the cooperative no longer increase in cash on hand. Combined, this means that $s^c_t$ is a piece-wise function of cash on hand and current milk production. Third, for levels of cash on hand between $A_1^*$ and $A_2^*$, it can be shown that the level of savings with the cooperative increases nonlinearly in cash on hand.\(^\text{10}\) In the empirical analyses, we will need to take into consideration these nonlinearities.

These predictions further imply that savings do not depend solely on cash on hand, $A_t$, but also on the market value of milk, $m_t$. This means that an exogenous increase in non-dairy income, $y_t$, will have different effects on the amount saved with the cooperative depending on the market value

\(^{10}\)To derive this result, we differentiate the FOC in (A.2) with respect to $A_t$. Savings are a nonlinear function of cash on hand under the assumption that the second derivative of utility is not a constant.
of milk production. When the market value of milk production is relatively high, it is likely that
\( A_t \leq A_2^*(m_t) \), so that savings with the cooperative are increasing in non-dairy income. When the
market value is low and hence \( A_2^*(m_t) \) is relatively low, it is more likely that \( A_t > A_2^*(m_t) \), and as a
result, savings with the cooperative cannot further increase in non-dairy income. We will therefore
study the effects of non-dairy and dairy income separately instead of using an aggregate for cash
on hand.\(^{11}\)

Another reason for studying the effects of these two variables individually is that they might
not be perfectly fungible. Households may engage in mental income accounting, or use the co-
operative as a commitment savings device for their dairy income, without having a similar device
for non-dairy income (cf. Casaburi and Macchiavello, 2016; Kramer and Kunst, 2017). Relatedly,
deviating from existing habits to regularly save one type of income can be mentally costly. This
would result in a differential treatment of dairy versus non-dairy income. Also, different household
members may control income from non-dairy activities versus dairy income. In the study context,
men are typically known to control payments from the cooperative, while women’s main sources
of income often include the proceeds from selling vegetables in the market as a source of non-dairy
income. We will hence allow dairy and non-dairy income to have different effects on recurring
savings deposits.

\(^{11}\)In the appendix, we show that savings in the low-return asset, \( s^f \), are zero until cash on hand reaches a third
threshold, \( A_3^*(m_t) > A_2^*(m_t) \). Above this threshold, these savings are increasing nonlinearly in cash on hand. We are
however mainly interested in the effects of cash on hand on recurring savings deposits, which we can empirically
observe from dairy sales to the cooperative. It is more challenging to obtain reliable empirical measures of recurring
savings in cash and other liquid assets. We therefore focus on sales to the cooperative and do not further discuss the
relation between cash on hand and buffer stock savings.
3 Methods

3.1 Context

We test our theoretical predictions empirically using high-detail high-frequency panel data collected year-long for a sample of 120 dairy farmers and their families. All farmers were members of Tanykina, a dairy cooperative in Nandi County near Eldoret in western Kenya, where dairy farming is an important income source for a large proportion of the population.\footnote{Formally, Tanykina 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 Tanykina or the cooperative to indicate Tanykina Dairies, Ltd.} Tanykina collects milk through several milk transporters and collection centers. At the start of the month, Tanykina agrees with larger milk processing companies on a price for milk collected from Tanykina that month, and the processor picks up all milk from Tanykina collection centers on a daily basis. Tanykina however pays farmers only once per month for milk delivered in the prior month, deferring milk payments by on average 25 days.\footnote{Tanykina may pay farmers on a monthly basis for various reasons. First, milk processors pay Tanykina only once per month, after assessing the actual quantity of milk collected in a given month. Once the processor has determined how much Tanykina will receive, the cooperative can start preparing farmers’ milk payments. In addition, paying on a monthly basis reduces transaction costs, and farmers may prefer monthly payments because this gives them a savings device.}

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 the costs of shares and any services utilized from Tanykina. Other than collectively marketing milk, Tanykina provides veterinary services and operates a store with agricultural inputs such as animal feed and fertilizers. These services are available to farmers unconditional on selling milk to Tanykina, but farmers who sell milk to Tanykina can pay for these through their milk account. The vast majority of purchases from Tanykina in the diaries (77.8 percent of 433 transactions) were indeed on credit. Although identifying consequences of side-selling for access to inputs is beyond the scope of this paper, it is
important to note that reducing the share of dairy sales to Tanykina could affect farmers’ ability to make these productivity-enhancing investments.

Further, through its Savings and Credit Cooperative Organization (SACCO), Tanykina provides agricultural loans and savings accounts. In practice, these services are not used very often. Throughout the year, the diaries recorded only two cases in which farmers took out a loan from the cooperative. For a similar sample of dairy farmers, Jack et al. (2016) also find a very low 2.4 percent take-up of SACCO loans, in part due to strict borrower requirements. Further, these loans are intended for agricultural investments, not for non-agricultural expenditures or to cope with financial shocks. Savings accounts do not bear interest and are mainly used to meet deposit requirements for accessing loans. The diaries recorded only 24 transactions in which a farmer deposited cash with the cooperative. Our empirical analyses hence abstract from these savings and credit provisions.

We will analyze how dairy savings are affected not only by cash on hand, but also by health shocks. At the time of the study, Tanykina members could enroll in a health insurance plan named The Community Health Plan (TCHP). TCHP was offered by AAR Insurance, a Kenyan insurance provider, with support from the PharmAccess Foundation and Health Insurance Fund. For members deciding to enroll their family into 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. For households who did not pay the monthly premium, insurance coverage was suspended the next month. To reinstate coverage, the household could pay a double premium in the next month; otherwise, the household was dropped from TCHP for the rest of the year. The program provided financial protection from health shocks and thereby improved consumption smoothing, in particular among households with weaker social networks (Geng et al., 2017).

---

14 Initially the program targeted Tanykina members and their families but the program has transitioned into a program that is accessible for the general public.
3.2 Data collection

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 understanding of health-seeking behavior and financial lives in the TCHP target population. The diaries project randomly selected seven villages with a minimum of 25 Tanykina member households from three dairy collection areas. From these seven villages, 120 randomly selected Tanykina member households were randomly selected to participate in the study. The number of selected Tanykina members sampled within a village was proportional to the total number of Tanykina members in that village, stratified by insurance status.

For these 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. Enumerators started each interview by collecting—at the household level—information about family members’ health, milk production and milk consumption, followed by interviews about respondents’ individual financial transactions since the last interview, including 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’; e.g. whether a payment was in cash, via mobile money or on credit), and the transaction value (‘how much’). We will use these records to identify the weekly value of milk sold to the cooperative versus other buyers in the market. Because milk production is recorded at the household level, we will aggregate the individual responses regarding financial transactions into household-level values.

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 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. However, the individual financial transactions for absent household members are missing in those 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 into the TCHP program. We consider a household to have insurance coverage if and only if at least one family member has (non-suspended) insurance coverage.

3.3 Econometric strategy: Semiparametric model and estimation

Our conceptual framework suggests that the relation between cash on hand and savings with the cooperative is nonlinear with threshold effects. We therefore estimate a semiparametric model, originally proposed by Su and Ullah (2006). This model includes both a nonparametric part for cash on hand and a parametric (linear) part to control for other covariates and household fixed effects:

$$S_{it} = \alpha_i + f(m_{it-1}, y_{it-1}) + x_{it}' \beta + v_{it}, \quad i = 1, \ldots, n, \quad t = 1, \ldots, T,$$

(3)

where $S_{it}$ is an indicator for the share of dairy sales to the cooperative for household $i$ in week $t$. Further, $\alpha_i$ is a household-specific effect to control for unobserved heterogeneity in household characteristics; $f(\cdot)$ is an unknown smooth function of past milk production ($m_{it-1}$) and past non-dairy income ($y_{it-1}$); $x_{it}$ is a $p \times 1$-vector of controls, including the log of current milk production ($m_t$), the median milk price in the village and its lag (to control for price changes in local market),
household consumption ($c_t$), health shocks, insurance coverage and their interaction (as alternative sources of variation in cash needs), and center-month fixed effects (to control for unobserved spatial variation over time); and $v_{it}$ represents unobserved time-varying heterogeneity in the share of dairy income saved with the cooperative.

Since $f(\cdot)$ is an unknown smooth function, past milk production and lagged non-dairy income—our two main determinants of cash on hand—enter the model nonparametrically. Our model predicts that the effect of cash on hand on savings with the cooperative will be nonlinear and characterized by threshold effects. Further, these two variables will not be perfect substitutes; they can have different effects because the farmer cannot increase savings with the cooperative above the value of milk production, and because the two sources might not be fungible. A fully linear model could include the two variables separately along with an interaction term to model this non-fungibility, but the individual effects would be presumed to be linear in the other variable and sensitive to the level at which the other variable would be centered.

The advantage of a semiparametric model is that it is not subject to this constraint. The nonparametric part allows for a specification-free, point-wise and data-driven estimator for the effect of cash on hand, while the linear part avoids a curse of dimensionality when controlling for potential confounds including unobservable household fixed effects. As such, a semiparametric model can yield a better description of how the variables included in the nonparametric part affect savings with the cooperative. This, in turn, increases the precision of the estimated coefficients for variables included in the linear part.

We estimate Equation (3) using the profile likelihood method, specifically, profile least squares, which builds on the following idea. If the linear parameters were given, we could rearrange Equation (3) as $S_{it} - \alpha_i - x_{it}'\beta = f(m_{it-1}, y_{it-1}) + v_{it}$ and obtain an explicit expression for the estimator of the nonparametric part, with $S_{it} - \alpha_i - x_{it}'\beta$ as the new dependent variable. This estimator is however infeasible since the linear parameters, $\alpha_i$ and $\beta$, are unknown. In the original estimating equation, we can nonetheless substitute $f(\cdot)$ with the infeasible nonparametric estimator and
rearrange to obtain parametric estimators using traditional ordinary least squares. The feasible
nonparametric estimator, given the parametric one, then follows immediately. The appendix
describes this estimation process more formally and provides details regarding implementation. We
refer readers to Su and Ullah (2006) for detailed proofs.

Estimator consistency requires that $E(v_{it}|x_i, m_i, y_i, \alpha_i) = E(v_{it}|x_{it}, m_{it-1}, y_{it-1}) = 0$, where $x_i = (x_{i1}, \cdots, x_{iT})'$, and where $m_i$ and $y_i$ are analogously defined. This means that error terms are i.i.d.
To ensure that this assumption is valid in practice, we include past instead of current realizations of non-dairy income and dairy production in the nonparametric part of the equation. Financial shocks could influence current realizations through coping strategies such as increased labor supply or forgoing investments in agriculture or business. Past realizations have the advantage of being predetermined, making the i.i.d. assumption more plausible, while still influencing current cash on hand through unobserved buffer stock savings, $s^d_{it-1}$, and through savings from past milk sales to the cooperative, $s^c_{it-1}$. In theory, both types of savings are a nonlinear function of past milk production, $m_{it-1}$, and of past non-dairy income, $y_{it-1}$.

We will consider the effects of these two variables on three different outcomes. Our main
outcome is the share of dairy sales to Tanykina ($\text{ShareSalesTan}$), as a proxy for the share of dairy
sales for which farmers defer payment. To construct this variable, we divide the transaction value
of milk sold to Tanykina ($SalesTan$) by total milk sales ($SalesTotal$) in a given week:

$$\text{ShareSalesTan} \equiv \frac{SalesTan}{SalesTotal} \equiv \frac{(1 + r^c)s^c_t}{m_t + r^c s^c_t}.$$

In order to obtain a closer proxy of the amount saved relative to the market value of milk,
$s^c_t/m_t$, we will also analyze the share of milk sold to the cooperative. We do not observe this

---

15 For identification, we assume $\sum_{i=1}^n \alpha_i = 0$. Because the sum of the $n$ household fixed effects is zero, it cannot capture
the sample mean. This will be captured instead by the nonparametric part $f(\cdot)$ and other coefficients in the linear part
of Equation (3).

16 Because savings do not only depend on cash on hand, $A_{ih}$, but also on the value of current milk production, $m_{it}$, we
still control for current milk production and current market prices.
variable directly, because the questionnaires recorded the value of every transaction instead of the number of items or the price per item sold. We can however estimate it as follows. First, we divide earnings from the cooperative, $SalesTan$, by its fixed monthly milk price, $PriceTan$, to obtain the quantity of milk sold to the cooperative. Second, we divide this by the total quantity of milk sold, $MilkSold$ (measured as milk production minus milk consumption). This yields the share of milk sold to Tanykina, $ShareMilkTan$:

$$ShareMilkTan \equiv \frac{SalesTan/PriceTan}{MilkSold} \equiv \frac{s_c^t}{m_t}.$$ 

We will also test our prediction that an increase in cash on hand raises the average price at which a farmer sells milk, $PriceAve$. Under the assumption that the mark-up offered by the cooperative is constant and positive, $r^c > 0$, an increase in cash on hand—increasing the share of milk sold to Tanykina—increases a farmer’s earnings per liter of milk sold. Empirically, the cooperative offers a positive mark-up, but this mark-up is not constant over time and could be related to either cash on hand or the share of milk sold in the market. It is hence important to test this prediction and assess whether an increase in cash on hand—potentially increasing savings with the cooperative—results in higher earnings per liter of milk sold. We estimate this variable as the ratio of total sales, $SalesTotal$, and the quantity of milk sold, $MilkSold$:

$$PriceAve \equiv \frac{SalesTotal}{MilkSold} \equiv \frac{m_t + r^c s_c^t}{m_t}.$$ 

---

Note that under the assumption of constant returns on saving with the cooperative, this variable is determined only by the share of milk sold to Tanykina and the share of dairy sales to Tanykina, since it can be written as follows:

$$PriceAve = \frac{ShareMilkTan}{ShareSalesTan} (1 + r^c) = \frac{s_c^t}{m_t} \frac{m_t + r^c s_c^t}{(1 + r^c)s_c^t} (1 + r^c) = \frac{m_t + r^c s_c^t}{m_t}.$$
3.4 Descriptive statistics

Table 1 presents an overview of household characteristics at the start of the diaries data collection, 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. The analyses are hence restricted to the 88 households with variation in the share of dairy sales to Tanykina described in Columns (1) and (2).

[Table 1 about here.]

In 70.5 percent of these households, the main decision-maker is male. The average age of the household head is 52.4. A common pattern among Kenyan dairy farmers is that men sell milk produced in the morning to the cooperative, while women sell milk produced in the afternoon in the market. Consistent with this, approximately half of all households have at least two different members who report selling milk at some point during the year. The average household has 4.2 cows, while at most one in three households is keeping bulls. Households have around 4.5 calves, of which the majority are female as well.

Only one-fifth of dairy farmers are male, and in half of the sample, the main household member selling milk is not the (male) household head but his spouse. She is on average 48 years of age, which is five years below the average age of 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 reports being engaged in livestock activities, but not all of them report being engaged in cattle activities (which mainly involves dairy farming in the study area) at that time. If engaged in cattle activities, 86.1 percent keeps some of the income from those activities for him- or herself, and a comparable percentage can decide on how to spend that money.

\(^{18}\) We can however identify a main dairy farmer, who reports on average 90.2 percent of all dairy sales within the household.
Columns (3) and (4) describe households without variation in the share of dairy sales to Tanykina. These households either sell always to the cooperative, or always in the market, and we compare them with the analysis sample in Columns (5) and (6). Households without any variation in where they sell their milk are more likely to have a female household head, and fewer household members sell milk within these households. Moreover, they have on average one cow less at baseline. The main dairy farmer within these households is significantly more likely to be the household head, and less likely to be married. This is consistent with the observation that these are often female-headed households, which will include widowed women.

Figure 1 presents percentage histograms for our three outcome variables, focusing on the analysis sample of households with variation in the share of dairy sales to Tanykina. We only include weeks in which the household sells milk. In more than 30 percent of those weeks, they report zero dairy sales to Tanykina, and they report dairy sales exclusively to Tanykina nearly 35 percent of the time (see the top panel). In other interview weeks, sales to Tanykina vary mostly between 30 and 55 percent of total dairy sales. This could be because households sell milk produced in the morning to Tanykina, while milk produced in the afternoon is sold in the market, or because households sell to Tanykina on some days, while selling to market vendors on other days within the same week.

The middle panel provides a percentage histogram for the share of milk sold to Tanykina. We estimate the quantity of milk sold to Tanykina in a given week as the reported sales to Tanykina divided by the monthly Tanykina price, and we divide this variable by the total quantity of milk sold. Because sales to Tanykina and the total quantity of milk sold are reported in different sections of the diaries, we introduce measurement error. As a result, the estimated quantity of milk sold to Tanykina can be larger than the total quantity of milk sold, and we censor these observations to one. Further, when the household reports no sales to other buyers, the estimated quantity of milk sold to Tanykina can still be smaller than the total quantity of milk sold. As a result, we do not
observe clustering at one. This measurement error underscores why we use the share of dairy sales to Tanykina as our preferred outcome variable.

The bottom panel provides a histogram of the average price at which farmers sell their milk, constructed by dividing total dairy sales by the quantity of milk sold. Farmers sell their milk at an average price that ranges between 20 and 35 Kenyan Shillings (KSh) per liter of milk.\footnote{The value of 1,000 KSh was approximately 11.50 USD at the time of data collection.} In this histogram, we have truncated this variable, meaning that the 26 (0.1 percent) observations that take on a value greater than 60 are not included in the histogram. When using this variable as an outcome in the analyses, we will be using the log price so that these few outliers do not influence our results.

Table 2 summarizes statistics for these and other time-variant variables that we will be using in the analyses. Columns (1)–(3) focus on households with variation in the share of dairy sales to Tanykina. Column (4) presents the difference in means between this analysis sample and the sample of households without variation. Column (5) presents the \( p \)-value from a \( t \)-test for differences in means, clustering standard errors at the household level.

[Table 2 about here.]

Panel A summarizes our three outcome variables. In weeks that households sell milk, total dairy sales are 1,572 KSh per week. Half of these sales are generated by transactions with Tanykina (\( \text{ShareSalesTan} \)), which defers payments; other sales are generated by transactions in the local market, generally with payment upon delivery.\footnote{One of the key differences between Tanykina and buyers in the market is that the former defers payments until the next month, while the latter tend to pay cash immediately. Other buyers defer payments in 211 out of 2,242 (9.4 percent) interview weeks. As a result, the share of dairy sales on credit is closely correlated with the share of dairy sales to Tanykina (\( \rho = 0.96 \)). Appendix Figure C2 shows that using as dependent variable the share of dairy sales on credit (as opposed to Tanykina) does not affect our results. The figure estimates Equation (3) for the share of dairy sales to be received on credit, i.e., the share of dairy sales for which the farmer defers payments. Appendix Figure C2b draws the slopes with respect to past milk production and non-dairy income. Findings are very comparable to those in Figures 2. In weeks with less cash on hand, households receive relatively less dairy income not only from Tanykina, but also from other buyers who defer payments.}
Households sell on average 53 liters of milk per week and 39 percent of this milk is sold to Tanykina (ShareMilkTan). The share of dairy sales to Tanykina is higher than the share of milk sold to Tanykina because Tanykina pays a higher price than vendors in the local market. Tanykina pays on average 36.2 KSh per liter, which is 19.3 percent higher than the average price of 29.2 KSh that farmers earn per liter of milk sold (PriceAve).

We do not observe market prices at the transaction level, but we did collect consumer market prices through a monthly survey with three vendors. In this survey, the average milk price was 30.52 KSh per liter during the study period. Since farmers receive a lower milk price than the price at which vendors sell their milk, this is likely an upper bound for the actual average market price. Appendix Figure C1 draws the monthly milk price offered by Tanykina and these milk prices observed in local markets by month. In every month, Tanykina prices are above the market prices, meaning a strictly positive mark-up $r^c > 0$. The mark-up however varies between 9.4 and 32.9 percent, potentially hiding even larger fluctuations across individual farmers, which are potentially correlated with cash on hand. This is why we include PriceAve as one of our outcome variables, and why we control for the median market price at which farmers from the same village sell their milk. Monthly averages for this variable are drawn in the figure as well.

Panel B summarizes our main explanatory variables. We proxy cash on hand by past income from non-dairy activities and past milk production, $y_{it-1}$ and $m_{it-1}$, respectively. To construct the former, we take the log ratio of non-dairy cash income and non-food expenditures (both in 1,000 KSh) in the previous week ($L1NonDairyInc$). Average non-dairy household income in cash is on average around 2,000 KSh (23 USD) per week, non-food expenditures are on average 2,500 KSh (29 USD), and their ratio is on average 0.917. We use the ratio of these two variables because 70.2 percent of non-food expenditures are related to agriculture and business, making $y_{it-1}$ a proxy for non-dairy income net of expenditures incurred while generating that income. Moreover, controlling

\footnote{Casaburi and Macchiavello (2016) and Kramer and Kunst (2017) do not replicate this result, perhaps because farmers in their setting have more access to traders that sell outside the local market.}
for household fixed effects, we observe a strong correlation between non-dairy income and non-food expenditures ($\rho = 0.49$), meaning that non-dairy income relative to non-food expenditures is a better proxy for cash on hand than non-dairy income itself.$^{22}$

For past milk production, $m_{it-1}$, we use log average milk production (in liters per week) in the last calendar month ($LmMilkProd$). The average household produces 72 liters of milk per week. Abstracting from milk consumption and sales to other buyers, this variable is a proxy for past milk sales to the cooperative, and hence for current payments from the cooperative, given that the cooperative pays farmers only once per month for all milk delivered in the previous month. We use this variable instead of last month’s dairy sales to the cooperative because the two variables are closely correlated ($\rho = 0.65$), while the inclusion of past dairy sales would mean including a lagged dependent variable. Including the lag of an endogenous choice variable would potentially introduce serial correlation in the residuals, violating our identification assumption of i.i.d. residuals.

As an alternative measure of cash needs, we will also test whether health problems affect milk selling decisions. In around one quarter of all weeks, at least one household member faces a health problem. Although half of the sample had insurance coverage for health expenditures at the start of the diaries, many households dropped out during the year, so that on average, households were insured in about one third of all weeks. This means that we can estimate the effect of health shocks for households without and with health insurance coverage. In the face of health problems within the household, those without health insurance will need cash in order to pay their medical bills. Our hypothesis is hence that health problems—especially those without health insurance coverage—increase the share of dairy sales to vendors in the market, and reduces the share of dairy

$^{22}$We exclude food expenditures from this ratio because we interpret it as the household consumption variable from our conceptual framework, $c_t$. The lagged food expenditures variable is hence the lag of an endogenous choice variable and including it would be the equivalent of including a lagged dependent variable in a model with household fixed effects. This would introduce a Nickell bias. We do not consider non-food expenditures as such an endogenous choice variable since unavoidable spending on health care and school fees constitute a large share of non-food expenditures net of expenditures on agriculture and business. Within households, the correlation between food expenditures and non-dairy income ($\rho = 0.27$) is also significantly lower than the correlation between non-food expenditures and non-diary income ($\rho < 0.01$).
sales to Tanykina. We will analyze the effect of health problems in the current or past week, given that households may not immediately seek health care, and that even if they seek health care in the same week, they may need to pay their medical bills in the following week.

Panel C presents other variables that we use as controls in the analyses. We estimate the village median of the market milk price, on average 26.7 KSh, as follows. First, we estimate the quantity of milk sold in the market by subtracting from the milk quantity sold our estimated quantity sold to Tanykina. Second, we divide the transaction value of dairy sales to buyers in the market by this estimated quantity. We then take the median at the village-week level, instead of using individual-level prices, to minimize measurement error and to ensure that weeks in which the farmer sells exclusively to Tanykina can be included in the analyses. We also control for food expenditures, which are on average 537 KSh per week, or approximately one-fifth of non-food expenditures.

We apply a log transformation to income and expenditure variables as well as milk production and milk prices. This serves to limit the influence of outliers, as the distribution for these variables is often skewed to the right. The estimated coefficients can hence be interpreted as the effect of a one percent increase in the variable shown. Our analyses will be conditional on the household selling milk. We observe dairy sales in 85 percent of all weeks, with milk being fully consumed by the household in other weeks. Column (3) presents the standard deviation of these variables within households. Although there is more variation across households, the variables included to estimate the model vary considerably within households, justifying the use of a household fixed effects approach.

Columns (4)-(5) compare the analysis sample with the sample of households for whom we observe no variation in the share of dairy sales to Tanykina. In comparison with these households, households in our analysis sample sell significantly larger quantities of milk per week, but they

---

23 In total, around one third of milk production is used for household consumption, and conditional on selling milk, the household consumes about one quarter of milk production. Purchasing milk is rare; households buy milk in only 0.4 percent of all weeks.
also produce significantly more milk. As a result, the average share of milk sold to Tanykina is not significantly higher. Furthermore, the analysis sample has significantly lower non-dairy income relative to non-food expenditures. As a result, dairy farming makes up a higher share of household income. These households are also more likely to sell milk in a given week, consistent with the observation that they have more cows and produce larger quantities of milk. It is important to keep this in mind when interpreting our results.

4 Results

4.1 Fully linear model

We first provide a fully linear approximation to Equation (3). To capture potential interaction effects in $f(\cdot)$, this linear model includes not only the main effects but also a linear interaction term for our two measures of cash on hand, past milk production ($m_{it-1}$) and non-dairy income ($y_{it-1}$):

$$S_{it} = \alpha_i + m_{it-1} \gamma_1 + y_{it-1} \gamma_2 + m_{it-1} \times y_{it-1} \gamma_3 + x_{it} \beta + v_{it}, \quad i = 1, \ldots, n, \quad t = 1, \ldots, T. \quad (4)$$

This specification is not based on theoretical economic foundations, but is a good candidate for a fully parametric model of how different sources of cash on hand affect farmers’ preferences as to where to sell their milk. We will later analyze to what extent our semiparametric estimates overlap with the fully linear estimates presented here.

Table 3 estimates Equation (4), controlling for household fixed effects $\alpha_i$, and as controls in $x_{it}$ we include current milk production $m_{it}$ (in logs), a variable to indicate the incidence of health problems in weeks $t$ and $t-1$ (equal to zero in case of no health problems, equal to one half if there were health problems in either week $t$ or week $t-1$, and equal to one if there were health problems in both weeks), the average level of insurance coverage in the last two weeks (defined
similarly), an interaction of the two, food expenditures as a proxy for household consumption \( c_t \), the village-week median for the price of milk sold in the market and its one-week lag (both in logs), and center-month fixed effects, which will capture variation in the monthly cooperative milk prices as well as other seasonal center-wide effects.

[Table 3 about here.]

Given the inclusion of an interaction term, estimates for \( \gamma_1 \) and \( \gamma_2 \)—the individual effects of past milk production, \( m_{it-1} \), and non-dairy income, \( y_{it-1} \), respectively—will depend on the level at which these two variables are centered. Panels A, B and C of Table 3 therefore estimate these two coefficients after centering past milk production and non-dairy income at the 25th, 50th and 75th percentile, respectively. Estimates for the interaction term are insensitive to where we center these two variables, and are hence presented only once.

Column (1) estimates the effects of last month’s milk production and last week’s non-dairy income on \( ShareSalesTan \), the share of dairy sales to Tanykina (as opposed to buyers in the market). In all three panels, an increase in past milk production is associated with a large significant increase in the share of dairy sales to Tanykina. If past production increases by one percent, the share of sales to Tanykina increases by 5.3 percentage points for the median farmer in Panel B \( (p < 0.01) \). Lagged non-dairy income does not have a significant effect on this outcome variable, but the interaction of past milk production and lagged non-dairy income is negative and statistically significant \( (p < 0.10) \). In other words, as households have less cash on hand from past non-dairy activities, an increase in past milk production has a stronger effect on the share of dairy sales that is generated through transactions with Tanykina.

Column (2) provides estimates of the fully linear model for the share of milk sold to Tanykina. Unlike the share of dairy sales to Tanykina, this variable is measured indirectly, but it corresponds conceptually with the share of current dairy income saved, \( s_c/m_t \). As in Column (1), past milk production has a positive and statistically significant effect, independent of whether the variables
are centered at the 25th, 50th or 75th percentile. For the median farmer, a one-percent increase in last month’s milk production is associated with a 6.5 percentage points increase in the share of milk sold to Tanykina \((p < 0.01\), see Panel B). Last week’s non-dairy income does not influence the share of milk sold to Tanykina, but the interaction term with last month’s milk production is negative as in Column (1).

In Appendix Table C1, we show that an increase in last month’s milk production is associated both with a significant decrease (increase) in the probability of selling no (all) milk to Tanykina. Last week’s non-dairy income ratio has no significant effect. Thus, on average, an increase in cash inflows from past dairy sales increases the probability of selling milk to Tanykina, and reduces the chances that farmers engage in side-selling. Income from other activities appears to have no effect on this decision.

Our conceptual framework predicts that an increase in the share of milk sold to the cooperative is associated with an increase in income. This prediction rests on the assumption that the mark-up offered by the cooperative is a positive constant, \(r_c > 0\). Because this might not be true empirically, Column (3) presents estimates for the average milk price at which a farmer sells milk. An increase in past milk production—increasing the share of milk sold to Tanykina, which pays higher prices than the market—is associated with higher earnings per liter of milk sold, consistent with our theoretical predictions. A ten-percent increase in past milk production raises the average price that a farmer obtains for his or her milk by 0.43 percent, which is only weakly significant \((p < 0.10)\). Lagged non-dairy income does not have a statistically significant effect on average prices, except for a small negative effect when both past milk production and lagged non-dairy income are evaluated at the 75th percentile.\(^{24}\)

\(^{24}\) We are focusing on the average price, that is, earnings per liter of milk sold, instead of total dairy sales, because dairy sales also depend on the quantity of milk sold, which by itself can be related to past milk production and non-dairy income. In Appendix Table C2, we show that past milk production indeed increases the probability of selling milk—see Column (1)—as well as the share of milk sold, both unconditional and conditional on selling milk—see Columns (2) and (3), respectively. Focusing on total dairy sales would hence yield stronger effects of past milk production, especially when also including interview weeks in which the household did not report any milk sales.
Appendix Table C3 analyzes whether our main results in the fully linear model are robust to including milk production in the next month and non-dairy income in the next week. The idea is that our key findings could be driven by comovements between the share of dairy sales to Tanykina and milk production instead of a causal effect of past milk production (and cash inflows from the cooperative) on the subsequent share of dairy sales generated through transactions with Tanykina. In that case, our outcome variables should be correlated not only with past realizations of milk production, but also with future realizations. We find that past milk production remains correlated significantly with the share of dairy sales to Tanykina and the share of milk sold to Taynkina, whereas future milk production does not bear a significant coefficient. This supports our hypothesis that increasing cash on hand from dairy farming affects the willingness to defer income. In Column (3), we show that this does not improve the average price at which the farmer sells milk.\footnote{As another robustness check, we control for potential comovements by assessing whether findings are robust to using household fixed effects over shorter periods of time. Table C4 estimates the fully linear model with household fixed effects estimated separately for the first half of the study year (the dry season) and the second half of the year (the wet season). Of course, by increasing the number of household-specific fixed effects, we are reducing substantially the amount of variation within households and precision, but we find comparable patterns to the ones estimated in Table 3.}

Thus, based on the fully linear model, we would conclude that an increase in past milk production is associated with higher savings rates of dairy income. Because the cooperative offers higher prices than the market, meaning that the return on these savings is positive, this also raises farmers’ average sales price, although this finding is not very robust and statistically significant only at the 10 percent level. An increase in non-dairy income does not influence the decision to defer milk payments.

4.2 Semiparametric estimates

This section compares these fully linear estimates with estimates of the semiparametric model in Equation (3). We will first focus on the nonparametric part for past milk production and past non-dairy income, \( f(m_{it-1}, y_{it-1}) \), followed by a discussion of the coefficients estimated for variables
4.2.1 Nonparametric part

To visualize the joint effects of past milk production and non-dairy income on the share of dairy sales to Tanykina, Figure 2a presents a three-dimensional (3D) graph, plotting the fitted share of dairy sales to Tanykina, \( \hat{f}(\cdot) \), against last month’s milk production and last week’s non-dairy income. These fitted values control for household fixed effects and variables included in \( x_{it} \).\(^{26}\) In the right-hand figure, we compare these fitted values with the predicted values based on the fully linear estimates. The share of dairy sales to Tanykina is increasing in milk production, which is consistent with the findings from the fully linear model. Moreover, there appears to be a positive effect of lagged non-dairy income on the share of dairy sales to Tanykina for a wide range of past milk production levels. The estimates of the fully linear model did not yield this result.

[Figure 2 about here.]

In order to give a closer examination of the individual effects of these two variables, and compare the semiparametric estimates with those from the fully linear model, Figure 2b presents the fitted slopes of the share of dairy sales to Tanykina, \( \hat{\dot{f}}(\cdot) \). We draw these slopes with respect to last month’s milk production (top panel), and last week’s non-dairy income (bottom panel), together with 95% confidence intervals. The figure evaluates these slopes at different levels of last month’s milk production (varied on the horizontal axis, with vertical dotted lines indicating its 25%, 50% and 75% quantile) and last week’s non-dairy income (at the 25%, 50% and 75% quantile in the left, middle and right figures, respectively). For comparison, we also compare these slopes with the estimates from the fully linear model, which imposes the condition that slopes with respect to past

\(^{26}\) For ease of comparison across specifications, values are rescaled to have the same mean as the raw dependent variable. The plot is derived using point-wise estimates of \( f(\cdot) \) and its slopes at 100 × 100 evenly spaced points across the 10%-90% quantile of past milk production and non-dairy income.
milk production (non-dairy income) are constant in past milk production (non-dairy income).27

The top panel draws slopes with respect to last month’s milk production to indicate the effect of this variable on the share of dairy sales to Tanykina. At the 25th percentile of lagged non-dairy income (the left figure), we observe a significant positive slope only at higher levels of past milk production, with the most pronounced effects at levels just above the median milk production level in the last month. Evaluated at the 50th percentile of lagged non-dairy income (the middle figure), the slope is significantly positive for a wide range of past milk production levels across the x-axis, although past milk production has more pronounced effects at lower (in particular below-median) quantiles. At the 75th percentile of lagged non-dairy income (the right figure), the slope is significantly positive only for below-median values of past milk production.

Thus, also in the semiparametric model, an increase in past milk production is associated with a higher share of dairy income saved with the cooperative, but the slopes depend on levels of both past milk production and past non-dairy income; increased levels of cash on hand from past milk production induces higher savings only when either past milk production is relatively high and past non-dairy income is low (the left figure), or when past non-dairy income is higher (the middle and right figures) but past milk production is relatively low. This is consistent with the prediction of nonlinearities and threshold effects, and the fully linear model fails to reveal this heterogeneity.

The fully linear model further estimates that an increase in lagged non-dairy income does not affect the share of dairy sales to Tanykina. The bottom panel of Figure 2b presents semiparametric estimates of this effect. At lower levels of lagged non-dairy income (left figure), slopes are close to zero and statistically insignificant, consistent with the conclusion from the fully linear model. At the median level of non-dairy income (middle figure), we observe a significant and positive effect for a wide range of past milk production levels, especially between the 25th and 50th percentile of

---

27 This means that in the top panel, these estimates can vary from the left to the right, but not along the x-axis within the graph; whereas in the bottom panel, these estimates can vary along the x-axis within the graph, but are identical for the left, middle and right figures.
past milk production. In the right figure, we observe even more pronounced effects. Thus, lagged non-dairy income does have a positive effect on the share of dairy sales to Tanykina. The fully linear model yields a different result because the effect is nonlinear. Further, the effects occur only at higher levels of non-dairy income combined with relatively lower levels of past milk production, consistent with the notion of threshold effects.

Next, we estimate Equation (3) for the proportion of milk sold to Tanykina, $ShareMilkTan$. Figure 3a presents a 3D graph for the fitted share of milk sold to Tanykina ($f(\cdot)$), and Figure 3b draws slopes with respect to last month’s milk production (top panel), and the non-dairy income ratio (bottom panel). Patterns in Figures 3 for the share of milk sold to Tanykina are very similar to those drawn in Figures 2 for the share of dairy sales to Tanykina. Households sell a relatively larger share of milk to Tanykina in weeks following a period of higher milk production and higher income from non-dairy activities; in other words, they save more with the cooperative in weeks that they have more cash on hand from either dairy farming or from non-dairy activities.

However, as in Figure 2, we only observe effects of past milk production when either lagged non-dairy income is low and past milk production is relatively high, or when lagged non-dairy income is higher and past milk production is relatively low. Moreover, when lagged non-dairy income is higher and past milk production is relatively low, we observe effects of the non-dairy income variable itself. Hence, also for the share of milk sold to Tanykina, we find threshold effects and nonlinearities in the effects of past milk production and lagged non-dairy income, as well as a nonlinear interaction effect. This is why the semiparametric estimates lead to different conclusions than the fully linear estimates.

As a final analysis, Figures 4 present fitted levels and slopes for the average price at which farmers sell their milk in a given week. Note that we control for the median price at which farmers in the same village sell milk, as well as center-month fixed effects, which capture variation in the
Tanykina price that is fixed on a monthly basis. In the top panel of Figure 4b, the average price at which households sell their milk is increasing in cash on hand from past milk production at lower levels of past milk production. This is in line with the finding that also the share of milk sold to Tanykina increases in cash on hand following a period of lower milk production, at least when evaluated at the 50th or 75th percentile of lagged non-dairy income, translating into higher average prices.

At the same time, at higher levels of past milk production, an increase in past milk production is associated with a reduction in the prices that farmers receive on average, and in the bottom panel of Figure 4b, the average price is decreasing in non-dairy income. In other words, a reduction in cash on hand can be associated with an increase in the average price at which households sell their milk. Since a reduction in cash on hand from non-dairy activities also reduces the share of milk sold to Tanykina, and since we control for the median market price for farmers in the same village, as well as center-month effects that capture the monthly Tanykina price, this increase in price must mean that a farmer receives a relatively higher market price than others in weeks when he or she has less non-dairy income.

One explanation for this finding is that there could be economies of scale in selling in the informal market. As a reduction in cash on hand from non-dairy activities induces farmers to sell more milk to market vendors (as opposed to Tanykina), they may receive a better price per liter of milk due to fixed costs of selling to a vendor. Alternatively, perhaps these buyers provide informal insurance when income from non-dairy activities declines, either unknowingly, for instance by purchasing diluted milk, or knowingly, by providing extra cash when farmers ask for it.\(^\text{28}\) It could also be that the effects of past milk production are generated by dairy farmers who are the main

\(^{28}\) Enumerators should have recorded these transactions separately as a gift or loan, but may have included in the amount received from the buyer.
recipient of Tanykina milk payments—typically male farmers or household heads—and that the effects of lagged non-dairy incomes are generated by dairy farmers who have less control over Tanykina milk payments, in particular female dairy farmers in male-headed households. We will further explore gender heterogeneity as a potential mechanism in Section 4.3.

In sum, our main findings indicate that farmers increase their dairy sales to the market (relative to the cooperative) in weeks that follow periods of lower milk production and non-dairy income, i.e., weeks in which a household has less cash on hand. Regarding past milk production, we would have reached the same conclusion based on the estimates of the fully linear model. However, those estimates would have led to the conclusion that lagged non-dairy income does not influence the share of milk sold to Tanykina. The fully linear estimates provide an incomplete—and sometimes even inaccurate—summary of how cash on hand is related to the decision to defer milk payments due to threshold effects and nonlinearities. Moreover, unlike the fully linear model, the semiparametric estimates reveal heterogeneity in how cash on hand affects the average sales price.

Interpretation of the differences between the fully linear and semiparametric results is straightforward but important. The linear model allows the non-dairy income effect to vary only in log milk production, not in its own level, and only linearly. In other words, the effect of a one-percent increase in non-dairy income is assumed to be constant across the entire range of non-dairy income levels. Moreover, it is assumed to have a constant effect on the slope with respect to past production across the entire range of past production levels. The nonparametric estimator allows this effect to vary in both production and non-dairy income in a nonlinear manner. As a result, it is largest for higher levels of non-dairy income, especially when past milk production is at median levels. The linear specification averages out these local effects, biasing the estimates and inference.

### 4.2.2 Fully parametric (linear) part

Table 4 provides the estimates for variables that are included as controls in the linear part of the semiparametric model. As an alternative proxy for cash needs, we analyze how households respond
to health shocks. In Column (1), health symptoms do not significantly influence the share of dairy sales to Tanykina. However, focusing on the share of milk sold to Tanykina, we do find a significant effect of health shocks ($p < 0.10$). Experiencing a health symptom in both the current and the last week decreases the share of milk sold to Tanykina by 3.1 percentage points if the household is uninsured.

[Table 4 about here.]

It is very likely that in weeks with health shocks, households sell more milk in the market in order to obtain cash to pay their medical bills. Consistent with this interpretation, we do not observe a negative effect of past health symptoms on both the share of dairy sales to Tanykina, and the share of milk sold to Tanykina, for households with health insurance coverage; the interaction term of health insurance coverage and health symptoms is significantly positive and outweighs the negative effect of health symptoms in Columns (1) and (2). Thus, insurance coverage reduces households’ tendency to sell their milk in the market in order to cope with health problems.

We also control for food expenditures to proxy household consumption, $c_t$. This variable is endogenous: farmers may increase food consumption when they have more cash on hand, which corresponds to weeks in which they are inclined to sell relatively more milk to Tanykina. Alternatively, an increase in food consumption may induce farmers to sell more milk to market vendors in order to obtain cash. We find more evidence of the former hypothesis. The share of dairy sales to Tanykina is increasing in food expenditures ($p < 0.10$), although we do not observe significant effects on our other outcome variables.

We also control for the median price at which farmers in the same village sell their milk. Higher prices are associated naturally with a higher average milk price in Column (3), but at the same time, the share of milk sold to Tanykina is significantly higher in weeks with a higher market price in Column (2). This could be related to reverse causality, that is, in weeks that farmers sell (ceteris paribus) more milk to Tanykina, there will be less milk available in the local market,
increasing the local market prices. The increase in market prices reduces the return on savings with Tanykina, meaning that in Column (1) the share of dairy income received from Tanykina—which is inflated by the return on savings with Tanykina—is not affected by the median market price.

The lagged village market price is predetermined and can be interpreted as an alternative proxy for cash on hand. If the market price was higher in the past week, farmers will have more cash on hand from selling milk in the market that week. This variable indeed has a positive effect on the share of dairy sales to Tanykina, and on the share of milk sold to Tanykina. We observe a positive but statistically insignificant effect on the average milk price that a farmer receives in Column (3).

Lastly, we control for current milk production, \( m_{it} \). In Column (1), the share of dairy sales to Tanykina is increasing in current milk production. This is not driven by an increase in the share of milk sold to Tanykina in Column (2), but by a decrease in the average price at which the farmer sells in the market, reducing the average price that the farmer receives in Column (3). When farmers produce more milk, they receive on average lower prices for their milk in the market, causing an increase in the share of dairy sales to Tanykina.\(^{29}\)

### 4.3 Heterogeneity by gender, time of the month, and season

In this section, we explore different sources of heterogeneity, including gender, proximity to the milk payment, and season. In the local (Nandi) culture, cooperatives register the household head as the member of the cooperative. This means that women, who are often responsible for dairy farming activities, can only be registered with the cooperative if they are also the household head, which only happens if they are widowed or if their husbands have migrated to work elsewhere, for instance in a larger city. As a result, female dairy farmers in male-headed households may not

\(^{29}\) Note that we observe this negative correlation between milk production and the price received on average controlling for center-month fixed effects, and controlling for the median price at which milk is sold in the market in the village both that same week and the previous week. The negative correlation is hence unlikely driven by an aggregate supply-side effect on milk prices.
have full control over the monthly milk payment from the cooperative. They will instead rely more on market vendors and on other sources of income from non-dairy activities. In this section, we therefore estimate our semiparametric model separately for two types of households: households with either a male dairy farmer or a female household head, versus households with a female dairy farmer and a male household head.30

Figure C3 presents estimated slopes of the share of milk sold to Tanykina for households with a male dairy farmer or a female household head in Panel (a), and for male-headed households with a female dairy farmer in Panel (b). In both panels, the figures in the top row draw slopes with respect to past milk production, as proxy for cash inflows from past dairy sales. Because the household head controls the cooperative account, female farmers in male-headed households will have less control over the monthly milk payment. This potentially explains why milk production in the past month affects the share of milk sold to Tanykina to a larger extent in Panel (a) than in Panel (b). The figures in the bottom rows draw slopes with respect to past non-dairy income, and reveal no strong heterogeneity: the share of milk sold to Tanykina increases in past non-dairy income when evaluated at median or above-median levels of lagged non-dairy income, and at below-median levels of past milk production.

Heterogeneity by gender could also offer a potential explanation for why we observe a negative effect, especially of past non-dairy income, on the average price at which a household sells milk. To explore this in more detail, Figure C4 presents the estimated slopes of the average price for milk sold when including only households with a male dairy farmer or a female head in Panel (a), and households with a female dairy farmer and a male head in Panel (b). The former group of households, which increases the share of milk sold to the cooperative in weeks with more cash on hand from past milk production, benefits from the higher cooperative prices in these weeks, as shown by the positive slopes in the top row of Panel (a) at below-median levels of past milk

30 In one female-headed household, the main dairy farmer is male. In other female-headed households, the main dairy farmer is female as well.
production. Female dairy farmers in male-headed households on the other hand, do not appear to capitalize on this savings opportunity in Panel (b). In fact, these dairy farmers are driving the negative effects of cash on hand from both past milk production and non-dairy income on average prices.

Next, we analyze whether the share of dairy sales from transactions with Tanykina responds differently to cash on hand depending on the time of the month. We expect that cash on hand from an increase in last month’s milk production will have a stronger effect for interviews in the second and third week of the month than in the first and last week, since the monthly Tanykina milk payment is due in the second week, and this milk payment will be increasing in the amount of milk delivered to the cooperative in the previous month. Figure C5 therefore presents slopes of the share of dairy sales to Tanykina estimated using interviews in the first and fourth week of the month in Panel (a), versus slopes estimated using interviews from the second and third week of the month in Panel (b). Consistent with our priors, we find the strongest effects of past milk production in the second and third week of the month (see the top row of figures in both panels). Moreover, we find no substantial heterogeneity in the effect of lagged non-dairy income (see the bottom row of figures in both panels). 31

As a final source of heterogeneity, we study whether effects of cash on hand are different in the first half of the study year (which was the dry season with lower milk production) versus the second half of the year (which was the wet season with higher milk production). Figure C6 draws slopes of the share of dairy sales to Tanykina with respect to past milk production and lagged non-dairy income estimated using interviews in the dry versus wet season in Panels (a) and (b), respectively. In both panels, the top row of figures draws the slopes with respect to past milk production. In the dry season, with lower milk production, an increase in past milk production affects the share of

31 We do not find heterogeneity when disaggregating the data by the first half of the month versus the last half of the month. That is, the number of days by which a farmer defers the milk payment when selling to Tanykina does not influence behavior.
sales to Tanykina at below-median levels of past milk production. In the wet season, with higher milk production, we observe similar but more pronounced effects. The effect of lagged non-dairy income, drawn in the bottom row of figures for both subsamples, has an effect only during the wet season in Panel (b).

These findings underscore the value of yearlong financial diaries; with a shorter panel, we would have been unable to identify such heterogeneity within the time span of a year, and we could have reached different conclusions regarding the role of cash constraints for saving in cooperatives versus side-selling. Moreover, because the diaries were collected every week, we can observe variation in households’ responses to an increase in cash on hand at different times within the same month. Finally, despite the relatively small number of households that were interviewed, we were able to identify interesting patterns of heterogeneity by gender. The diaries do not include sufficient detail on dairy farmers’ relationships with different types of buyers to explain why we observe this heterogeneity, but these findings do merit further research on the relations between different types of dairy farmers and market vendors.

5 Conclusion

This paper analyzed to what extent cash on hand affects the recurring decision to sell milk to a cooperative, which defers payments, versus buyers in the market, who generally pay cash upon delivery. We focus on a sample of Kenyan dairy farmers. In comparable settings, Casaburi and Macchiavello (2016) and Kramer and Kunst (2017) find that dairy farmer cooperatives function as a commitment savings device: farmers value the savings function of deferred payments from the cooperative and are even willing to sacrifice dairy income in order to be paid at a later date. We show that farmers nonetheless earn a significant share of their dairy income by selling milk in the market, especially when they are more cash-constrained, and even when buyers in the market offer
lower prices than the cooperative.

These analyses generalize the finding that cash on hand influences savings in lab-in-the-field experiments (Carvalho et al., 2016; Janssens et al., 2017) to recurring savings with a cooperative observed in a non-experimental setting. A priori, we expected these decisions to be of a more habitual nature, and subject to social norms and bylaws imposed by the cooperative agreement to deliver milk on a recurring basis (Casaburi and Macchiavello, 2015). As such, we did not necessarily expect these savings to be sensitive to variation in cash on hand. However, our findings indicate that this is not the case; cash on hand, as measured through past realizations of milk production and non-dairy income, has meaningful effects on the share of milk sold to the cooperative.

The present study has inevitably also a number of limitations. First, the diaries project provides highly detailed information on respondents’ financial transactions, but it was not designed with the intention to study milk marketing behavior. As a result, respondents reported the total transaction value by type of buyer in a given week, but did not report the milk price or quantity sold for that transaction. In addition, the diaries did not collect detailed information about respondents’ relationships with different buyers in the local market. This is a challenge in interpreting the effects on earnings per liter of milk sold. Understanding why we observe a negative effect of cash on hand on these average prices when restricting the sample to female dairy farmers in male-headed households remains an area for future research.

Second, our measures of cash on hand—past milk production and past non-dairy income—were not varied at random, and we do not introduce an exogenous policy change. We try minimizing concerns around endogeneity as much as possible by exploiting the panel nature of our data, which allow us to control for unobserved household fixed effects and to use lagged instead of contemporaneous values of these explanatory variables. Moreover, the linear part of the semi-parametric model controls for observable confounds. However, we do not test whether a policy change, for instance the introduction of cash transfers, cash advances or insurance mechanisms, help stabilize farmers’ milk selling patterns. In informing policy, it is important to replicate our
findings in alternative settings, potentially by introducing random variation in cash on hand, and to rigorously evaluate the effects of potential interventions that can help relax cash constraints.

Third, we interpret milk delivered to the cooperative as a recurring deposit. However, there are potentially other differences between buyers in the market and the cooperative that are not directly related to savings decisions. It could be that households turn to buyers in the market when they are more cash constrained not only because they pay upon delivery, but also because they allow farmers to sell lower-quality milk (introducing opportunities for diluting the milk), and because these buyers might provide informal gifts or loans. Qualitative interviews suggest that a farmer would never dilute his milk (there is a superstition that this will curse the cow) but households did report in these interviews that local market vendors may provide some extra cash. This could explain why we observe a negative effect of increasing cash on hand for the average price at which the farmer sells milk, and further research could explore these mechanisms in more detail.

We nonetheless believe that our findings have important implications. From a method perspective, the paper shows the potential value of estimating a semiparametric model when outcome variables are a nonlinear function of two different variables, in our case, two different sources of cash on hand. Based on a fully linear model, we would have concluded that only past milk production (not past non-dairy income) affects the share of milk sold to the cooperative, and without affecting the average price at which a farmer sells milk. The semiparametric estimates offer a more nuanced picture, showing an effect of past milk production as well as lagged non-dairy income. The two variables affect savings nonlinearly and in different ways, as an indication that they are not perfectly fungible. Such nonlinearities can also explain why we find interesting heterogeneity in the effect on average prices, which the fully linear model would have overlooked.

From a policy perspective, our findings provide a new explanation for the commonly observed high levels of side-selling—and accompanying variability in recurring savings—that undermine both households’ ability to save and the sustainability of many cooperatives. Cash-constrained households will be pushed towards side-selling even if the cooperative offers higher prices than
buyers in the local market, potentially lowering total dairy income. Access to financial instruments that can relax cash constraints, for instance cash advances or health insurance, potentially reduces the need for side-selling and improve the stability of recurring savings with the cooperative. This can have positive implications for dairy farmers’ ability to meet their savings goals, and the improved stability of milk supply may help strengthen cooperative performance and improve members’ livelihoods.

References


**Figure 1:** Percentage histograms for dairy marketing decisions

- **Top left diagram:** Share of dairy sales to Tanykina vs. Percentage [%]
- **Middle left diagram:** Share of milk sold to Tanykina vs. Percentage [%]
- **Bottom diagram:** Average milk price vs. Percentage [%]
Figure 2: Fitted share of dairy sales to Tanykina against lagged milk production and non-dairy income

(a) Nonparametric estimates (left) and stacked with the fully linear estimates (right)

(b) Fitted slopes of the share of dairy sales to Tanykina and 95% confidence intervals
Figure 3: Fitted share of milk sold to Tanykina against lagged milk production and non-dairy income

(a) Nonparametric estimates (left) and stacked with the fully linear estimates (right)

(b) Fitted slopes of the share of milk sold to Tanykina and 95% confidence intervals

<table>
<thead>
<tr>
<th>Slope at 25% L1NonDairyInc</th>
<th>Slope at 50% L1NonDairyInc</th>
<th>Slope at 75% L1NonDairyInc</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

slope at 25%, 50%, and 75% L1NonDairyInc  95% confidence interval (CI)  linear estimate and 95% CI
**Figure 4**: Fitted price received on average for milk sold against lagged milk production and non-dairy income

(a) Nonparametric estimates (left) and stacked with the fully linear estimates (right)

(b) Fitted slopes of average price of milk sold and 95% confidence intervals
<table>
<thead>
<tr>
<th></th>
<th>Variation in share of dairy sales to Tanykina</th>
<th>No variation in share of dairy sales</th>
<th>Difference in means (1) - (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (1)</td>
<td>s.d. (2)</td>
<td>Mean (3)</td>
</tr>
<tr>
<td>Household head is male</td>
<td>0.705</td>
<td>0.459</td>
<td>0.500</td>
</tr>
<tr>
<td>Age of the household head</td>
<td>52.38</td>
<td>14.15</td>
<td>51.03</td>
</tr>
<tr>
<td>Number of HH members selling milk</td>
<td>1.489</td>
<td>0.547</td>
<td>1.300</td>
</tr>
<tr>
<td>Number of cows at baseline</td>
<td>4.227</td>
<td>2.509</td>
<td>3.200</td>
</tr>
<tr>
<td>Number of cows lent out at baseline</td>
<td>0.125</td>
<td>0.333</td>
<td>0.069</td>
</tr>
<tr>
<td>Number of bulls at baseline</td>
<td>0.356</td>
<td>0.849</td>
<td>0.333</td>
</tr>
<tr>
<td>Number of male calves at baseline</td>
<td>1.227</td>
<td>1.444</td>
<td>0.833</td>
</tr>
<tr>
<td>Number of female calves at baseline</td>
<td>3.318</td>
<td>10.22</td>
<td>1.233</td>
</tr>
<tr>
<td>Number of weeks with an interview</td>
<td>45.42</td>
<td>4.085</td>
<td>46.03</td>
</tr>
</tbody>
</table>

**Main dairy farmer:**

<table>
<thead>
<tr>
<th></th>
<th>Variation in share of dairy sales to Tanykina</th>
<th>No variation in share of dairy sales</th>
<th>Difference in means (1) - (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (1)</td>
<td>s.d. (2)</td>
<td>Mean (3)</td>
</tr>
<tr>
<td>Is male</td>
<td>0.216</td>
<td>0.414</td>
<td>0.300</td>
</tr>
<tr>
<td>Age</td>
<td>47.57</td>
<td>14.34</td>
<td>46.40</td>
</tr>
<tr>
<td>Is household head</td>
<td>0.477</td>
<td>0.502</td>
<td>0.733</td>
</tr>
<tr>
<td>Is spouse of household head</td>
<td>0.500</td>
<td>0.503</td>
<td>0.200</td>
</tr>
<tr>
<td>Is married</td>
<td>0.830</td>
<td>0.378</td>
<td>0.600</td>
</tr>
<tr>
<td>Went to secondary school or higher</td>
<td>0.531</td>
<td>0.502</td>
<td>0.423</td>
</tr>
<tr>
<td>Is protestant</td>
<td>0.830</td>
<td>0.378</td>
<td>0.933</td>
</tr>
<tr>
<td>Is engaged in livestock activities</td>
<td>0.953</td>
<td>0.213</td>
<td>0.967</td>
</tr>
<tr>
<td>Is engaged in cattle activities</td>
<td>0.765</td>
<td>0.427</td>
<td>0.833</td>
</tr>
<tr>
<td>Is engaged in other livestock activities</td>
<td>0.635</td>
<td>0.484</td>
<td>0.567</td>
</tr>
<tr>
<td>Can keep part of cattle income</td>
<td>0.659</td>
<td>0.477</td>
<td>0.793</td>
</tr>
<tr>
<td>Decides how to spend cattle income</td>
<td>0.655</td>
<td>0.478</td>
<td>0.793</td>
</tr>
</tbody>
</table>

**Number of households**: 88 30

**Notes**: 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 (2 weeks), Easter (1 week) and the last fieldwork month (4 weeks). The main dairy farmer is the household member reporting the highest value of dairy income received throughout the year.
Table 2: Summary statistics of time-varying characteristics

<table>
<thead>
<tr>
<th>Panel A: Outcome variables</th>
<th>Variation in share of dairy sales to Tanykina</th>
<th>Difference in means comp. to farmers without variation in share of dairy sales</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (1)</td>
<td>s.d. (2)</td>
</tr>
<tr>
<td>Total dairy sales in 1,000 KSh</td>
<td>1.572</td>
<td>0.936</td>
</tr>
<tr>
<td>Share of dairy sales to Tanykina</td>
<td>0.503</td>
<td>0.413</td>
</tr>
<tr>
<td>Liters of milk sold per week</td>
<td>53.36</td>
<td>33.84</td>
</tr>
<tr>
<td>Share of milk sold to Tanykina</td>
<td>0.390</td>
<td>0.331</td>
</tr>
<tr>
<td>Tanykina milk price* (monthly price in KSh)</td>
<td>36.19</td>
<td>2.217</td>
</tr>
<tr>
<td>Average price of milk sold</td>
<td>29.16</td>
<td>17.54</td>
</tr>
</tbody>
</table>

Panel B: Main explanatory variables

| Weekly non-dairy cash income in 1,000 KSh | 2.009 | 4.065 | 2.924 | -0.052 | 0.920 |
| Weekly non-food expenditures in 1,000 KSh | 2.493 | 4.457 | 3.754 | 0.542 | 0.287 |
| Weekly non-dairy income ratio | 0.917 | 1.913 | 1.769 | -1.355 | 0.003 |
| Liters of milk produced per week | 71.50 | 37.49 | 19.16 | 17.70 | 0.008 |
| Prop. of weeks with health problem | 0.263 | 0.440 | 0.391 | -0.004 | 0.924 |
| Prop. of weeks with insurance coverage | 0.344 | 0.475 | 0.245 | -0.056 | 0.538 |

Panel C: Other time-variant characteristics

| Weekly food expenditures in 1,000 KSh | 0.537 | 0.912 | 0.844 | -0.014 | 0.891 |
| Village median of estimated market price | 26.72 | 4.839 | 3.878 | -2.221 | 0.000 |
| Prop. of weeks with dairy sales | 0.847 | 0.360 | 0.276 | 0.150 | 0.013 |
| Number of households (total observations) | 88 (3997) |

Notes: 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 (2 weeks), Easter (1 week) and the last fieldwork month (4 weeks). * October milk prices were not collected and have been imputed by November milk prices.
Table 3: Fully parametric (linear) estimates for milk marketing decisions

<table>
<thead>
<tr>
<th>Panel A. Centered at 25% quantile</th>
<th>Share of dairy sales to Tanykina</th>
<th>Share of milk sold to Tanykina</th>
<th>Log price received on average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Run milk production last month</td>
<td>0.060*** (0.021)</td>
<td>0.071*** (0.017)</td>
<td>0.048* (0.026)</td>
</tr>
<tr>
<td>Log non-dairy income ratio last week</td>
<td>0.015 (0.013)</td>
<td>0.021 (0.017)</td>
<td>-0.003 (0.016)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B. Centered at 50% quantile</th>
<th>Share of dairy sales to Tanykina</th>
<th>Share of milk sold to Tanykina</th>
<th>Log price received on average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Run milk production last month</td>
<td>0.053*** (0.020)</td>
<td>0.065*** (0.017)</td>
<td>0.043* (0.025)</td>
</tr>
<tr>
<td>Log non-dairy income ratio last week</td>
<td>0.003 (0.008)</td>
<td>0.012 (0.007)</td>
<td>-0.010 (0.011)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel C. Centered at 75% quantile</th>
<th>Share of dairy sales to Tanykina</th>
<th>Share of milk sold to Tanykina</th>
<th>Log price received on average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Run milk production last month</td>
<td>0.045** (0.021)</td>
<td>0.059*** (0.017)</td>
<td>0.038 (0.026)</td>
</tr>
<tr>
<td>Log non-dairy income ratio last week</td>
<td>-0.008 (0.009)</td>
<td>-0.007 (0.007)</td>
<td>-0.019* (0.011)</td>
</tr>
</tbody>
</table>

| Interaction term                  | -0.026* (0.015)                | -0.021* (0.013)                | -0.018 (0.019)                |

<table>
<thead>
<tr>
<th>Controls</th>
<th>Yes</th>
<th>Yes</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>R-squared within households</td>
<td>0.101</td>
<td>0.114</td>
<td>0.156</td>
</tr>
<tr>
<td>Mean dependent variable</td>
<td>0.498</td>
<td>0.384</td>
<td>3.307</td>
</tr>
<tr>
<td>Number of observations</td>
<td>2873</td>
<td>2873</td>
<td>2873</td>
</tr>
<tr>
<td>Number of households</td>
<td>88</td>
<td>88</td>
<td>88</td>
</tr>
</tbody>
</table>

Notes: The analysis is restricted to households with variation in the share of milk sold to Tanykina, and weeks in which these households sell milk. The log of variable $x$ is proxied by $\log(x + 1)$. Controls include current milk production (in logs), food expenditures in 1,000 KSh (in logs), the median market milk price in the village (current and first lag, both in logs), prop. health shocks in last two weeks (current and last week), prop. of last two weeks with insurance coverage, an interaction of these two variables, and center-month fixed effects. Standard errors in parentheses. $^* p < 0.10, ^{**} p < 0.05, ^{***} p < 0.01$. 

50
Table 4: Semiparametric estimates for milk marketing decisions (linear part)

<table>
<thead>
<tr>
<th></th>
<th>Share of dairy sales to Tanykina</th>
<th>Share of milk sold to Tanykina</th>
<th>Log price received on average</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>Health symptoms in household</td>
<td>-0.024</td>
<td>-0.031*</td>
<td>-0.017</td>
</tr>
<tr>
<td></td>
<td>(0.023)</td>
<td>(0.017)</td>
<td>(0.021)</td>
</tr>
<tr>
<td>Household has insurance coverage</td>
<td>0.013</td>
<td>0.012</td>
<td>0.105***</td>
</tr>
<tr>
<td></td>
<td>(0.026)</td>
<td>(0.020)</td>
<td>(0.036)</td>
</tr>
<tr>
<td>... X Health symptoms in household</td>
<td>0.057*</td>
<td>0.046**</td>
<td>-0.011</td>
</tr>
<tr>
<td></td>
<td>(0.030)</td>
<td>(0.023)</td>
<td>(0.037)</td>
</tr>
<tr>
<td>Log milk production this week</td>
<td>0.111***</td>
<td>0.013</td>
<td>-0.382**</td>
</tr>
<tr>
<td></td>
<td>(0.020)</td>
<td>(0.020)</td>
<td>(0.060)</td>
</tr>
<tr>
<td>Log food expenditures in 1,000 KSh</td>
<td>0.026*</td>
<td>0.020</td>
<td>-0.003</td>
</tr>
<tr>
<td></td>
<td>(0.014)</td>
<td>(0.013)</td>
<td>(0.030)</td>
</tr>
<tr>
<td>Village-week median log milk price in market</td>
<td>-0.031</td>
<td>0.048**</td>
<td>0.353***</td>
</tr>
<tr>
<td></td>
<td>(0.026)</td>
<td>(0.021)</td>
<td>(0.086)</td>
</tr>
<tr>
<td>... (lagged)</td>
<td>0.049**</td>
<td>0.061***</td>
<td>0.101</td>
</tr>
<tr>
<td></td>
<td>(0.024)</td>
<td>(0.021)</td>
<td>(0.070)</td>
</tr>
<tr>
<td>Center-month effects</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>R-squared within households</td>
<td>0.132</td>
<td>0.145</td>
<td>0.181</td>
</tr>
<tr>
<td>Mean dependent variable</td>
<td>0.498</td>
<td>0.407</td>
<td>3.316</td>
</tr>
<tr>
<td>Number of observations</td>
<td>2873</td>
<td>2873</td>
<td>2873</td>
</tr>
<tr>
<td>Number of households</td>
<td>88</td>
<td>88</td>
<td>88</td>
</tr>
</tbody>
</table>

Notes: The analysis is restricted to households with variation in the share of milk sold to Tanykina, and weeks in which these households sell milk. The log of variable $x$ is proxied by $\log(x+1)$. We report $R^2$ in the semiparametric regressions by the squared correlation coefficient between the demeaned $y$ (minus household mean across time) and fitted value of demeaned $y$, $\tilde{y}_i$: \[
\tilde{y}_i = y_i - \bar{y}_i = \left(x_i - \bar{x}_i\right)^\prime \hat{\beta} + \hat{m}(z_i) - \bar{\hat{m}}_i,
\]
where $\tilde{y}_i \equiv 1/T \sum_{t=1}^{T} y_{it}$, $\bar{x}_i \equiv 1/T \sum_{t=1}^{T} x_{it}$, and $\bar{\hat{m}}_i \equiv 1/T \sum_{t=1}^{T} \hat{m}(z_{it})$. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. 

51
Online Appendix

There are three parts in this appendix. In Part A, we provide the solution to the dynamic optimization problem put forward in our conceptual framework. In Part B, we provide a brief introduction to the local linear kernel estimator and give asymptotic properties of $\hat{\beta}$ and $\hat{M}(z)$, based on which we construct confidence intervals. We also describe consistent estimators of the covariance matrices and bandwidth selection method. Part C provides additional tables and figures that are not included in the main text.

A Solution to the dynamic optimization problem

In the dynamic consumption-saving model proposed in Section 2, the Bellman Equation $V(A_t) \equiv V(A_t, s_{t-1}^a, s_{t-1}^c, y_t, m_t)$ is given as

$$V(A_t) = \max_{s_{t}^a \geq 0, s_{t}^c \in [0, m_t]} \mathbb{E}_t \left\{ v(A_t - s_{t}^a - s_{t}^c) + \frac{1}{1+\delta} V(A_{t+1}) \right\}. \quad (A.1)$$

First-order conditions (FOCs) with respect to savings in the liquid asset, $s_{t}^a$, and savings from selling to the cooperative, $s_{t}^c$, respectively, are defined as follows,

$$-v'(c_t) + \frac{1+r^a}{1+\delta} \mathbb{E}_t v'(c_{t+1}) = 0 \quad \text{and} \quad -v'(c_t) + \frac{1+r^c}{1+\delta} \mathbb{E}_t v'(c_{t+1}) = 0, \quad (A.2)$$

whereby $\partial V(A_{t+1}) / \partial s_{t}^j = (1+r^j) \partial V(A_{t+1}) / \partial A_{t+1}$ for $j \in \{a, c\}$, $\partial V(A_{t+1}) / \partial A_{t+1} = v'(c_{t+1})$ by the Envelope Theorem, and $c_t = A_t - s_{t}^a - s_{t}^c$.  

32 Specifically, we apply the Envelope Theorem to the Bellman function $\partial V(A_t) / \partial A_t = v'(c_t)$, and iterate this expression forward one period, yielding $\partial V(A_{t+1}) / \partial A_{t+1} = v'(c_{t+1})$. 

52
Since farmers cannot borrow, $s_t^f \geq 0$ and $s_t^c \geq 0$, consumption cannot be higher than $A_t$, and both conditions will be violated if marginal utility from consuming $A_t$ is higher than discounted expected marginal utility from period $t + 1$. This can happen when farmers have bad draws of non-dairy income and milk production, as well as low savings from the previous period, so that $A_t$ takes on a low value, and the expressions in (A.2) will be strictly negative. Deaton (1991) shows that there exists a unique threshold $A^*$ such that when $A_t$ is below this critical level, we observe a corner solution with full consumption of cash on hand, while at higher levels, the FOC is no longer negative when evaluated at $c_t = A_t$, and hence it will be optimal to carry forward a positive savings balance in the liquid asset.

Within our framework, there are however two savings devices and hence two FOCs. Because the return on the liquid asset is smaller than the return on deferred milk payments from the cooperative, $r^a < r^c$, the marginal utility from saving in the liquid asset is strictly lower than the marginal utility from saving one’s dairy income with the cooperative. As a result, the farmer will always prefer saving with the cooperative. To implicitly define the unique threshold above which the farmer saves dairy income, we evaluate the FOC for that variable, $s_t^c$, at zero savings, $c_t = A_t$:

$$\frac{1 + r^c}{1 + \delta} E_t v'(c(y + m, m)) = v'(A^*_t)$$

whereby $c(A, m)$ with $A = y + m$ indicates the optimal consumption in the next period given cash on hand $A$ and milk production $m$ in that period, subject to zero savings in the current period.

At the same time, recall that farmers cannot save more than the market value of dairy production with the cooperative, $s_t^c \leq m_t$. At sufficiently high levels of cash on hand, the FOC will however be satisfied at a savings level that exceeds the market value of milk production. The threshold level of cash on hand at which this occurs depends on $m_t$, and is implicitly defined by evaluating the
FOC for $s_t^c$ at maximum savings with the cooperative, $s_t^c = m_t$:

$$\frac{1 + r^c}{1 + \delta} E_t v'(c(y + m + (1 + r^c)m_t, m)) = v'(A_t^2(m_t) - m_t)$$

whereby $c(A, m)$ with $A = y + m + (1 + r^c)m_t$ indicates the optimal consumption in the next period subject to the farmer selling all milk to the cooperative in the current period. For levels of cash on hand $A_t \geq A_t^2(m_t)$, we observe a second corner solution in which the farmer will choose to save his or her full dairy income with the cooperative, and additional savings demands will be met by saving in the liquid buffer stock asset.\(^{33}\) This yields the following prediction for the amount that is saved by selling to the cooperative:

**Lemma 1.** Consider the utility maximization problem given in (2) subject to (1). Let $c_t = c(A_t, m_t)$ be a stationary stochastic optimum. There exist unique thresholds $A_1^*$ and $A_2^*(m)$ with $A_2^*(m) > A_1^* > 0$ such that

1. $c = c(A, m) = A$ and $s^c = 0$, if $A \leq A_1^*$;
2. $c = c(A, m) = A - s^c$ and $0 < s^c < m$, if $A_1^* < A < A_2^*(m)$;
3. $c = c(A, m) \leq A - m$ and $s^c = m$, if $A \geq A_2^*(m)$.

**Proof.** First, assume that $A_t$ is low enough for the following condition to be satisfied:

$$\frac{1 + r^a}{1 + \delta} E_t v'(c_{t+1}) < \frac{1 + r^c}{1 + \delta} E_t v'(c_{t+1}) \leq v'(A_t). \quad (A.3)$$

\(^{33}\) When $A_t = A_2^*$, savings in the liquid asset will be zero because the marginal utility from saving in that asset is lower than the marginal utility from saving with the cooperative:

$$\frac{1 + r^a}{1 + \delta} E_t v'(c(y + m + (1 + r^c)m_t, m)) < \frac{1 + r^c}{1 + \delta} E_t v'(c(y + m + (1 + r^c)m_t, m)) = v'(A_2^*(m_t) - m_t)$$

Only once cash on hand exceeds a third threshold, we will observe savings in the liquid asset. Because our main interest is in analyzing empirically the amount that farmers save with the cooperative as a function of cash on hand, we do not include that third threshold in the remainder of the discussion.
In this case, the consumer will not save in either the liquid asset or with the cooperative; \( s^a_t = s^c_t = 0 \).

Second, consider the case in which \( A_t \) satisfies the following condition:

\[
\frac{1 + r^a}{1 + \delta} E_t v'(c_{t+1}) \leq v'(A_t) < \frac{1 + r^c}{1 + \delta} E_t v'(c_{t+1}) < v'(A_t - m_t). \tag{A.4}
\]

In this case, the consumer will not save in the liquid asset, \( s^a_t = 0 \), but will find an interior solution for the cooperative savings, \( s^c_t \in (0, m_t) \).

Third, consider the case in which \( A_t \) satisfies the following condition:

\[
\frac{1 + r^a}{1 + \delta} E_t v'(c_{t+1}) < v'(A_t - m_t) \leq \frac{1 + r^c}{1 + \delta} E_t v'(c_{t+1}). \tag{A.5}
\]

In this case, the consumer will not save in the liquid asset, \( s^a_t = 0 \), and will save as much as possible with the cooperative, \( s^c_t = m_t \).

Fourth, consider the case in which \( A_t \) satisfies the following condition:

\[
v'(A_t - m_t) < \frac{1 + r^a}{1 + \delta} E_t v'(c_{t+1}) < \frac{1 + r^c}{1 + \delta} E_t v'(c_{t+1}) \tag{A.6}
\]

Now, the consumer will save as much as possible with the cooperative, \( s^c_t = m_t \), and save any remaining savings demand in the liquid asset, \( s^a_t > 0 \), but at a lower return.

Combined, these four scenarios provide an augmented Euler equation for \( s^c_t \) and \( s^a_t \):

\[
v'(A_t - s^c_t) = \max \left[ v'(A_t), \min \left[ v'(A_t - m_t), \beta^c E_t v'(c_{t+1}) \right] \right],
\]

\[
v'(A_t - m_t - s^a_t) = \max \left[ v'(A_t - m_t), \beta^a E_t v'(c_{t+1}) \right], \tag{A.7}
\]

where \( \beta^a = (1 + r^a)/(1 + \delta) < 1 \) as \( r < \delta \) and \( \beta^c = (1 + r^c)/(1 + \delta) > \beta^a \).
B  Semiparametric estimator and its implementation

B.1  Local linear kernel estimator

To conceptualize this estimation process, let \( \alpha = (\alpha_2, \cdots, \alpha_n)' \), \( D = (I_n \otimes i_T) d_n \), \( d_n = (-i_{n-1}, I_{n-1})' \) and \( z \) is a vector with past milk production, \( m_{t-1} \), and net income, \( y_{it} \). We can then write Equation (3) in matrix form:

\[
Y = D\alpha + M + X\beta + v, \tag{B.1}
\]

where \( Y = (y_{11}, \cdots, y_{1T}, y_{21}, \cdots, y_{nT})' \), \( X = (x_{11}, \cdots, x_{1T}, x_{21}, \cdots, x_{nT})' \), \( v = (v_{11}, \cdots, v_{1T}, v_{21}, \cdots, v_{nT})' \), 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 \( \alpha \) and \( \beta \) were known, we have a purely non-parametric regression:

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

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

\[
M_{\alpha,\beta}(z) = S(z)(Y - D\alpha - X\beta), \quad m_{\alpha,\beta}(z) = s(z)'(Y - D\alpha - X\beta), \tag{B.2}
\]

where \( S(z) \) and \( s(z) \) are the local linear pre-multipliers, explicitly defined in the appendix.

3. Apply the unfeasible estimator \( m_{\alpha,\beta}(z_{it}) \) for \( m(z_{it}) \) in (B.1) and rearrange. By partitioned regression formula, we obtain the least square parametric estimators:

\[
\hat{\beta} = (X' M^* X^*)^{-1} X' M^* Y^*, \quad \hat{\alpha} = (\hat{\alpha}_2, \cdots, \hat{\alpha}_n)' = (D^* D^*)^{-1} D^* (Y^* - X^* \hat{\beta}). \tag{B.3}
\]

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 \( s_{it} \equiv s(z_{it}) \).
4. Plugging \( \hat{\beta} \) and \( \hat{\alpha} \) back to (B.2), we have the estimators for \( M(z) \) and \( m(z) \):

\[
\hat{M}(z) = S(z)(Y - D\hat{\alpha} - X\hat{\beta}), \quad \hat{m}(z) = s(z)'(Y - D\hat{\alpha} - X\hat{\beta}). \tag{B.4}
\]

Estimator for \( \alpha_1 \) follows as \( \hat{\alpha}_1 = -\sum_{i=2}^{n} \hat{\alpha}_i \).

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

\[
M_{\alpha,\beta}(z) = \arg \min_{\theta \in \mathbb{R}^3} \sum_{i=1}^{n} \sum_{t=1}^{T} \left( y_{it}^* - \theta_0 - (z_{it}' - z')(\theta_1, \theta_2) \right)^2 K(H^{-1}(z_{it} - z))
\]

\[
= \arg \min_{\theta \in \mathbb{R}^3} \left( Y^* - \tilde{Z}(z)\theta \right)'K_H(z)\left( Y^* - \tilde{Z}(z)\theta \right), \tag{B.5}
\]

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

Minimization in (B.5) gives \( M_{\alpha,\beta}(z) = (\tilde{Z}(z)'K_H(z)\tilde{Z}(z))^{-1}\tilde{Z}(z)'K_H(z)Y^* \equiv S(z)Y^* \), 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 pre-multipliers in Equation (B.2) and (B.4).
B.2 Asymptotic properties of $\hat{\beta}$ and $\hat{M}(z)$

Asymptotic properties of $\hat{\beta}$ and $\hat{M}(z)$ are established in Su and Ullah (2006) in Theorem 3.1 and 3.2 respectively under assumptions A1-A7. For $\hat{\beta}$, let $\bar{x}_{it} = x_{it} - E(x_{it} | z_{it})$, $\bar{x}_{it} = x_{it} - (s(z_{it})'X)'$, and $\bar{v}_{it} = y_{it} - x_{it}'\hat{\beta} - \hat{m}(z_{it}) - \hat{\alpha}_t$, we have

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} \mathcal{N}(0, \Sigma), \quad \tag{B.6}$$

where $\Sigma = \Phi^{-1}\Omega\Phi^{-1}$, $\Phi = \sum_i E(\bar{x}_{it} (\bar{x}_{it} - \bar{x}_{is} / T))$, and $\Omega = \sum_i \sum_r E(\bar{x}_{it} (\bar{x}_{is} - \bar{x}_{il} / T)'v_{it}v_{is})$. A consistent estimator for $\Sigma$ is given by $\hat{\Sigma} = \Phi^{-1}\hat{\Omega}\Phi^{-1}$, where $\Phi = n^{-1}\sum_i \sum_r \bar{x}_{it} (\bar{x}_{is} - \bar{x}_{il} / T)'\bar{v}_{it}\bar{v}_{is}$.

For $\hat{M}(z)$, let $\bar{f}(z) = \sum_{i=1}^{T} f_i(z)$, $\bar{v}_{it} = v_{it} - T^{-1}\sum_{s=1}^{T} v_{is}$, $\sigma^2_i(z) = E(v_{it}^2 | z_{it} = z)$, and $\delta^2(z) = \sum_{i=1}^{T} \sigma^2_i(z)f_i(z)$, we have

$$\sqrt{n[H]} \left( \hat{M}(z) - M(z) - Q^{-1} \begin{pmatrix} \frac{1}{2} tr(\int_{\mathbb{R}^2} uu'K(u)duH\hat{m}(z)H) \\ 0 \end{pmatrix} \right) \xrightarrow{d} \mathcal{N}(0, Q^{-1}\Gamma Q^{-1}), \quad \tag{B.7}$$

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

$$Q = \bar{f}(z) \begin{pmatrix} 1 & 0' \\ 0 & \int_{\mathbb{R}^2} uu'K(u)du \end{pmatrix}, \quad \Gamma = \delta^2(z) \begin{pmatrix} \int_{\mathbb{R}^2} K(u)^2du & 0' \\ 0 & \int_{\mathbb{R}^2} uu'K(u)du \end{pmatrix}.$$

The asymptotic normal distribution derived in Equation (B.7) can be used to obtain pointwise confidence intervals for the non-parametric estimator $\hat{M}(\cdot)$. As it is argued in 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, $\hat{m}(\cdot)$. Thus we choose a smaller bandwidth than 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 $k(u) = 0.75(1 - u^2)1(|u| \leq 1)$. Let
\[ \mu_i \equiv \int_\mathbb{R} u^i k(u) \, du, \quad v_i \equiv \int_\mathbb{R} u^i k(u)^2 \, du. \]  
Then, \[ f_{\mathbb{R}^2} uu' K(u) \, du = \begin{pmatrix} \mu_2 \mu_0 & \mu_2^2 \\ \mu_2^2 & \mu_2 \mu_0 \end{pmatrix} = \begin{pmatrix} 1/5 & 0 \\ 0 & 1/5 \end{pmatrix}, \]
and \[ \int_\mathbb{R}^2 K(u)^2 \, du = \nu_0^2 = 3/5. \]

By Equation (B.7), it is easy to obtain the element-wise asymptotic normal distribution for \( \hat{m}(\cdot) \) and the slope estimator, \( \hat{m}_i(\cdot) \), for \( i = 1, 2 \),

\[
\sqrt{n|H|} \left( \hat{m}(z) - m(z) - \frac{\mu_2}{2} \sum_{i=1}^2 h_i^2 \hat{m}_{ii}(z) \right) \xrightarrow{d} \mathcal{N} \left( 0, \frac{\nu_0^2 \sigma^2(z)}{f^2(z)} \right), \tag{B.8}
\]

\[
\sqrt{nh_i^2|H|} \left( \hat{m}_i(z) - \hat{m}_i(z) \right) \xrightarrow{d} \mathcal{N} \left( 0, \frac{\nu_0 v_i \hat{\sigma}^2(z)}{\mu_2^2 f^2(z)} \right), \tag{B.9}
\]

where \( \hat{f}(z) = \sum_{i} f_i(z) \), \( \hat{\sigma}^2(z) = \sum_{i=1}^T \sigma_i^2(z) f_i(z) \), \( \hat{m}_i(z) \) is the \( i \)th element in the the first derivative vector of \( m(\cdot) \) at \( z \), and \( \hat{m}_{ii}(z) \) is the \( i \)th diagonal element in the the second derivative matrix of \( m(\cdot) \) at \( z \).

A limited number of observations is a challenge in estimation the variance function for each period. For simplicity, we assume homoscedasticity and the same joint density function of the covariates across time, i.e., \( \sigma_i^2(z) = \sigma_f^2 < \infty \) and \( f_i(z) = f(z) \). Then, we have \( \hat{\sigma}^2(z) = \sum_{i=1}^T \hat{\sigma}_i^2(z) f_i(z) \). A consistent estimator for \( \sigma_i^2 \) is \( \hat{\sigma}_i^2 = n^{-1} \sum_{i=1}^n \hat{v}_i^2 \), where \( \hat{v}_i = \hat{v}_i - T^{-1} \sum_{i=1}^T \hat{v}_i \). We use the consistent Rosenblatt density estimator \( \hat{f}(z) \) to approximate \( f(z) \), where \( \hat{f}(z) \equiv (n|H|)^{-1} \sum_{i=1}^T \hat{f}_i(z) K(H^{-1}(z_{it} - z)) \). \( K(\cdot) \) is the Epanechnikov product kernel and \( H \) is obtained using cross-validation method.

Thus, based on (B.8), (B.9), and estimators for the covariances, the point-wise 95\% confidence intervals of \( \hat{m}(z) \) and \( \hat{m}_i(z) \) are constructed respectively as CI(z) = [\( \hat{m}(z) - 2s(z), \hat{m}(z) + 2s(z) \)], CI_i(z) = [\( \hat{m}_i(z) - 2s_i(z), \hat{m}_i(z) + 2s_i(z) \)], where \( s(z) = ((\hat{m}_1 \hat{h}_2 T^2 \hat{f}(z))^{-1} v_0^2 \sum_{i=1}^T \hat{\sigma}_i^2(z))^{1/2} \) and \( s_i(z) = ((\hat{m}_1 \hat{h}_2 T^2 \hat{f}(z))^{-1} v_0 v_i \sum_{i=1}^T \hat{\sigma}_i^2(z))^{1/2} \), for \( i = 1, 2 \).
In non-parametric estimation, bandwidth selection is crucial in order to obtain a good estimate. We choose bandwidths using 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 sample deviation of \( z_i \) for \( i = 1, 2 \). We choose \( h \) to be

\[
\hat{h} = \arg \min_{h \in \mathbf{H}} CV(h) \equiv \arg \min_{h \in \mathbf{H}} \left\{ \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \left( y_{it} - \hat{m}_{-it}(z_{it}) \right)^2 \right\}, \tag{B.10}
\]

where \( \hat{m}_{-it}(z_{it}) \) is the local linear estimator for \( \mathbb{E}(y_{it} | z_{it} = z) \) obtained by deleting the observation \( z_{it} \). Consistent with Su and Ullah (2006), \( \hat{h} \) 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. Denote the constants that associated with \( \hat{h} \) as \( (\hat{c}_1, \hat{c}_2) \). Note that there is an asymptotic bias term in \( \hat{M}(z) \), as is shown in (B.7). It’s more convenient to undersmooth a little bit (i.e., let \( nh^6 \to 0 \) as \( n \to \infty \)) than estimating the unknown term \( \hat{m}(z) \) 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 Rosenblatt density estimators, we choose cross-validation method that minimizes estimate of the integrated squared error (ISE), \( \int (\hat{f}(z) - f(z))^2 \, dz \). In particular,

\[
\hat{h}_d = \arg \min_{h} CV_d(h) \equiv \arg \min_{h} \left\{ \frac{1}{(nT)^2 |H|} \sum_{i_1=1}^{n} \sum_{i_2=1}^{T} \sum_{t_1=1}^{n} \sum_{t_2=1}^{T} \tilde{K} \left( H^{-1}(z_{i_1t_1} - z_{i_2t_2}) \right) \right. \\
- \left. \frac{2}{nT(nT-1)|H|} \sum_{i_1=1}^{n} \sum_{t_1=1}^{T} \sum_{(i_2,t_2) \neq (i_1,t_1)} K \left( H^{-1}(z_{i_1t_1} - z_{i_2t_2}) \right) \right\}, \tag{B.11}
\]

where \( \tilde{K}(\cdot) \) is the convolution function of the kernel \( K(\cdot) \). For Epanechnikov product kernel, \( \tilde{K}(u) = \tilde{k}(u_1)\tilde{k}(u_2) \), where \( \tilde{k}(u) = \int k(x)k(x-u) \, dx = \frac{3}{160}(2-|u|)^3(u^2 + 6|u| + 4)1(|u| \leq 2) \). \( \hat{h}_d \) 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 Nelder-Mead simplex direct search algorithm developed in Lagarias et al. (1998).
C Appendix Figures and Tables

Figure C1: Milk price paid by Tanykina versus average milk price in market
**Figure C2:** Fitted share of dairy sales for which the farmer defers payments

(a) Nonparametric estimates (left) and stacked with the fully linear estimates (right)

(b) Fitted slopes and 95% confidence intervals for share of income deferred

<table>
<thead>
<tr>
<th>Slope at 25% $L_{1NonDairyInc}$</th>
<th>Slope at 50% $L_{1NonDairyInc}$</th>
<th>Slope at 75% $L_{1NonDairyInc}$</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Graph" /></td>
<td><img src="image2.png" alt="Graph" /></td>
<td><img src="image3.png" alt="Graph" /></td>
</tr>
<tr>
<td><img src="image4.png" alt="Graph" /></td>
<td><img src="image5.png" alt="Graph" /></td>
<td><img src="image6.png" alt="Graph" /></td>
</tr>
</tbody>
</table>
Figure C3: Fitted slopes of the share of milk sold to Tanykina and 95% confidence intervals

(a) Households with male dairy farmer or female household head

Slope at 25% $L1_{\text{NonDairyInc}}$

Slope at 50% $L1_{\text{NonDairyInc}}$

Slope at 75% $L1_{\text{NonDairyInc}}$

(b) Households with female dairy farmer and male household head

Slope at 25% $L1_{\text{NonDairyInc}}$

Slope at 50% $L1_{\text{NonDairyInc}}$

Slope at 75% $L1_{\text{NonDairyInc}}$

- slope at 25%, 50%, and 75% $L1_{\text{NonDairyInc}}$
- 95% confidence interval (CI)
- linear estimate and 95% CI
Figure C4: Fitted slopes of average price of milk sold and 95% confidence intervals

(a) Households with male dairy farmer or female household head

Slope at 25% $L_{1}NonDairyInc$

Slope at 50% $L_{1}NonDairyInc$

Slope at 75% $L_{1}NonDairyInc$

(b) Households with female dairy farmer and male household head

Slope at 25% $L_{1}NonDairyInc$

Slope at 50% $L_{1}NonDairyInc$

Slope at 75% $L_{1}NonDairyInc$
**Figure C5:** Fitted share of dairy sales to Tanykina against lagged milk production and non-dairy income by time of the month

(a) Interviews in the first and fourth week of the month

- **Slope at 25% L1NonDairyInc**
- **Slope at 50% L1NonDairyInc**
- **Slope at 75% L1NonDairyInc**

(b) Interviews in the second and third week of the month

- **Slope at 25% L1NonDairyInc**
- **Slope at 50% L1NonDairyInc**
- **Slope at 75% L1NonDairyInc**

---

**Linear estimate and 95% CI**
Figure C6: Fitted share of dairy sales to Tanykina against lagged milk production and non-dairy income by time of the season

(a) Interviews in the dry season

Slope at 25% L1NonDairyInc

Slope at 50% L1NonDairyInc

Slope at 75% L1NonDairyInc

(b) Interviews in the wet season

Slope at 25% L1NonDairyInc

Slope at 50% L1NonDairyInc

Slope at 75% L1NonDairyInc
Table C1: Factors influencing where the household sells milk

<table>
<thead>
<tr>
<th></th>
<th>Sells no milk to Tanykina (1)</th>
<th>Sells some milk to Tanykina (2)</th>
<th>Sells all milk to Tanykina (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A. Centered at 25% quantile</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log milk production last month</td>
<td>-0.055**</td>
<td>0.010</td>
<td>0.045**</td>
</tr>
<tr>
<td></td>
<td>(0.022)</td>
<td>(0.016)</td>
<td>(0.022)</td>
</tr>
<tr>
<td>Log non-dairy income ratio last week</td>
<td>-0.014</td>
<td>0.004</td>
<td>0.011</td>
</tr>
<tr>
<td></td>
<td>(0.016)</td>
<td>(0.012)</td>
<td>(0.016)</td>
</tr>
<tr>
<td><strong>Panel B. Centered at 50% quantile</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log milk production last month</td>
<td>-0.049**</td>
<td>0.009</td>
<td>0.040*</td>
</tr>
<tr>
<td></td>
<td>(0.022)</td>
<td>(0.016)</td>
<td>(0.022)</td>
</tr>
<tr>
<td>Log non-dairy income ratio last week</td>
<td>-0.007</td>
<td>0.003</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td>(0.011)</td>
<td>(0.008)</td>
<td>(0.011)</td>
</tr>
<tr>
<td><strong>Panel C. Centered at 75% quantile</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log milk production last month</td>
<td>-0.042*</td>
<td>0.008</td>
<td>0.034</td>
</tr>
<tr>
<td></td>
<td>(0.023)</td>
<td>(0.017)</td>
<td>(0.023)</td>
</tr>
<tr>
<td>Log non-dairy income ratio last week</td>
<td>0.006</td>
<td>0.001</td>
<td>-0.008</td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
<td>(0.007)</td>
<td>(0.009)</td>
</tr>
<tr>
<td><strong>Interaction term</strong></td>
<td>0.021</td>
<td>-0.003</td>
<td>-0.018</td>
</tr>
<tr>
<td></td>
<td>(0.017)</td>
<td>(0.012)</td>
<td>(0.017)</td>
</tr>
<tr>
<td>Controls</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>R-squared within households</td>
<td>0.076</td>
<td>0.053</td>
<td>0.057</td>
</tr>
<tr>
<td>Mean dependent variable</td>
<td>0.316</td>
<td>0.349</td>
<td>0.335</td>
</tr>
<tr>
<td>Number of observations</td>
<td>2873</td>
<td>2873</td>
<td>2873</td>
</tr>
<tr>
<td>Number of households</td>
<td>88</td>
<td>88</td>
<td>88</td>
</tr>
</tbody>
</table>

Notes: The analysis is restricted to households with variation in the share of milk sold to Tanykina, and weeks in which these households sell milk. The log of variable \( x \) is proxied by \( \log(x + 1) \). Controls include current milk production (in logs), food expenditures in 1,000 KSh (in logs), median market milk price in the village (current and first lag, both in logs), prop. health shocks in last two weeks (current and last week), prop. of last two weeks with insurance coverage, an interaction of these two variables, and center-month fixed effects. Standard errors in parentheses. * \( p < 0.10 \), ** \( p < 0.05 \), *** \( p < 0.01 \).
**Table C2: Factors influencing the total quantity of milk sold**

<table>
<thead>
<tr>
<th>Panel A. Centered at 25% quantile</th>
<th>Sells milk (1)</th>
<th>Share of production sold (2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log milk production last month</td>
<td>0.058***</td>
<td>0.033***</td>
<td>0.018***</td>
</tr>
<tr>
<td></td>
<td>(0.021)</td>
<td>(0.007)</td>
<td>(0.005)</td>
</tr>
<tr>
<td>Log non-dairy income ratio last week</td>
<td>0.003</td>
<td>-0.002</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>(0.012)</td>
<td>(0.004)</td>
<td>(0.004)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B. Centered at 50% quantile</th>
<th>Sells milk (1)</th>
<th>Share of production sold (2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log milk production last month</td>
<td>0.058***</td>
<td>0.035***</td>
<td>0.018***</td>
</tr>
<tr>
<td></td>
<td>(0.020)</td>
<td>(0.007)</td>
<td>(0.005)</td>
</tr>
<tr>
<td>Log non-dairy income ratio last week</td>
<td>0.002</td>
<td>0.002</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
<td>(0.003)</td>
<td>(0.003)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel C. Centered at 75% quantile</th>
<th>Sells milk (1)</th>
<th>Share of production sold (2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log milk production last month</td>
<td>0.057***</td>
<td>0.038***</td>
<td>0.019***</td>
</tr>
<tr>
<td></td>
<td>(0.020)</td>
<td>(0.007)</td>
<td>(0.005)</td>
</tr>
<tr>
<td>Log non-dairy income ratio last week</td>
<td>0.001</td>
<td>0.006</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>(0.010)</td>
<td>(0.003)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>Interaction term</td>
<td>-0.003</td>
<td>0.008</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>(0.015)</td>
<td>(0.005)</td>
<td>(0.004)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Conditional on selling milk</th>
<th>No</th>
<th>No</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>R-squared within households</td>
<td>0.258</td>
<td>0.581</td>
<td>0.417</td>
</tr>
<tr>
<td>Mean dependent variable</td>
<td>0.854</td>
<td>0.694</td>
<td>0.733</td>
</tr>
<tr>
<td>Number of households</td>
<td>88</td>
<td>88</td>
<td>88</td>
</tr>
<tr>
<td>Number of observations</td>
<td>3365</td>
<td>3365</td>
<td>2873</td>
</tr>
</tbody>
</table>

**Notes:** The analysis is restricted to households with variation in the share of milk sold to Tanykina. The log of variable $x$ is proxied by log$(x + 1)$. Controls include current milk production (in logs), food expenditures in 1,000 KSh (in logs), median market milk price in the village (current and first lag, both in logs), prop. health shocks in last two weeks (current and last week), prop. of last two weeks with insurance coverage, an interaction of these two variables, and center-month fixed effects. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. 

69
### Table C3: Controlling for future realizations of milk production and non-dairy income

<table>
<thead>
<tr>
<th></th>
<th>Share of dairy sales to Tanykina</th>
<th>Share of milk sold to Tanykina</th>
<th>Log price received on average</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
</tbody>
</table>

#### Panel A. Centered at 25% quantile

- Log milk production last month: 0.070*** (0.022) 0.076*** (0.018) 0.028 (0.027)
- Log non-dairy income ratio last week: 0.015 (0.014) 0.007 (0.011) -0.023 (0.016)

#### Panel B. Centered at 50% quantile

- Log milk production last month: 0.062*** (0.022) 0.071*** (0.018) 0.030 (0.026)
- Log non-dairy income ratio last week: 0.003 (0.009) -0.000 (0.007) -0.020* (0.011)

#### Panel C. Centered at 75% quantile

- Log milk production last month: 0.053** (0.022) 0.066*** (0.018) 0.032 (0.026)
- Log non-dairy income ratio last week: -0.010 (0.009) -0.008 (0.008) -0.017 (0.011)
- Interaction term: -0.028* (0.016) -0.016 (0.014) 0.007 (0.020)
- Log milk production next month: -0.029 (0.028) -0.019 (0.024) 0.004 (0.034)
- Log non-dairy income ratio next week: 0.002 (0.008) -0.003 (0.007) -0.017* (0.010)

**Controls**
- Yes
- Yes
- Yes

**Notes:** The analysis is restricted to households with variation in the share of milk sold to Tanykina, and weeks in which these households sell milk. The log of variable \( x \) is proxied by \( \log(x + 1) \). Controls include current milk production (in logs), food expenditures in 1,000 KSh (in logs), the median market milk price in the village (current and first lag, both in logs), prop. health shocks in last two weeks (current and last week), prop. of last two weeks with insurance coverage, an interaction of these two variables, and center-month fixed effects. Standard errors in parentheses. * \( p < 0.10 \), ** \( p < 0.05 \), *** \( p < 0.01 \).
Table C4: Introducing household-season fixed effects

<table>
<thead>
<tr>
<th></th>
<th>Share of dairy sales to Tanykina (1)</th>
<th>Share of milk sold to Tanykina (2)</th>
<th>Log price received on average (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel A. Centered at 25% quantile</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log milk production last month</td>
<td>0.041* (0.023)</td>
<td>0.061*** (0.019)</td>
<td>0.061** (0.029)</td>
</tr>
<tr>
<td>Log non-dairy income ratio last week</td>
<td>0.022* (0.012)</td>
<td>0.019* (0.010)</td>
<td>0.015 (0.016)</td>
</tr>
<tr>
<td>Panel B. Centered at 50% quantile</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log milk production last month</td>
<td>0.032 (0.023)</td>
<td>0.054*** (0.019)</td>
<td>0.054* (0.029)</td>
</tr>
<tr>
<td>Log non-dairy income ratio last week</td>
<td>0.008 (0.008)</td>
<td>0.007 (0.007)</td>
<td>0.004 (0.010)</td>
</tr>
<tr>
<td>Panel C. Centered at 75% quantile</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log milk production last month</td>
<td>0.022 (0.023)</td>
<td>0.046** (0.019)</td>
<td>0.045 (0.029)</td>
</tr>
<tr>
<td>Log non-dairy income ratio last week</td>
<td>-0.007 (0.008)</td>
<td>-0.005 (0.007)</td>
<td>-0.009 (0.011)</td>
</tr>
<tr>
<td>Interaction term</td>
<td>-0.032** (0.015)</td>
<td>-0.026** (0.013)</td>
<td>-0.027 (0.019)</td>
</tr>
<tr>
<td>Household-season fixed effect</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Controls</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>R-squared within households</td>
<td>0.103</td>
<td>0.118</td>
<td>0.105</td>
</tr>
<tr>
<td>Mean dependent variable</td>
<td>0.498</td>
<td>0.384</td>
<td>3.307</td>
</tr>
<tr>
<td>Number of observations</td>
<td>2873</td>
<td>2873</td>
<td>2873</td>
</tr>
<tr>
<td>Number of households</td>
<td>88</td>
<td>88</td>
<td>88</td>
</tr>
</tbody>
</table>

Notes: The analysis is restricted to households with variation in the share of milk sold to Tanykina, and weeks in which these households sell milk. The log of variable $x$ is proxied by log($x + 1$). Controls include current milk production (in logs), food expenditures in 1,000 KSh (in logs), the median market milk price in the village (current and first lag, both in logs), prop. health shocks in last two weeks (current and last week), prop. of last two weeks with insurance coverage, an interaction of these two variables, and center-month fixed effects. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. 
Table C5: Semiparametric estimates by household head and farmer gender (linear part)

<table>
<thead>
<tr>
<th></th>
<th>Share of milk sold to Tanykina</th>
<th>Log price received on average</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male farmer or female HH head</td>
<td>Male head and female farmer</td>
</tr>
<tr>
<td>Health symptoms in household</td>
<td>-0.034 (0.024)</td>
<td>-0.012 (0.033)</td>
</tr>
<tr>
<td>Household has insurance coverage</td>
<td>-0.012 (0.023)</td>
<td>-0.008 (0.024)</td>
</tr>
<tr>
<td>... X Health symptoms in household</td>
<td>0.083** (0.035)</td>
<td>0.035 (0.037)</td>
</tr>
<tr>
<td>Log milk production this week</td>
<td>-0.007 (0.025)</td>
<td>-0.178** (0.072)</td>
</tr>
<tr>
<td>Log food expenditures in 1,000 Sh</td>
<td>0.011 (0.017)</td>
<td>0.013 (0.021)</td>
</tr>
<tr>
<td>Village-week median log milk price in market</td>
<td>0.065*** (0.022)</td>
<td>0.667*** (0.105)</td>
</tr>
<tr>
<td>... (lagged)</td>
<td>0.047* (0.024)</td>
<td>0.297*** (0.072)</td>
</tr>
<tr>
<td>Center-month effects</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>R-squared within households</td>
<td>0.148</td>
<td>0.057</td>
</tr>
<tr>
<td>Mean dependent variable</td>
<td>0.444</td>
<td>3.336</td>
</tr>
<tr>
<td>Number of observations</td>
<td>1450</td>
<td>1450</td>
</tr>
<tr>
<td>Number of households</td>
<td>44</td>
<td>44</td>
</tr>
</tbody>
</table>

Notes: The analysis is restricted to households with variation in the share of milk sold to Tanykina, and weeks in which these households sell milk. The log of variable \( x \) is proxied by \( \log(x + 1) \). We report \( R^2 \) in the semiparametric regressions by the squared correlation coefficient between the demeaned \( y \) (minus household mean across time) and fitted value of demeaned \( y \), \( \tilde{y}_t \) and \( \hat{y}_t \): \( \tilde{y}_t = y_t - \bar{y}_i \) and \( \hat{y}_t = (x_t - \bar{x}_i)\hat{\beta} + \hat{m}(z_t) - \hat{m}_i \), where \( \tilde{y}_i = 1/T \sum_{t=1}^T y_t \), \( \bar{x}_i = 1/T \sum_{t=1}^T x_t \), and \( \hat{m}_i = 1/T \sum_{t=1}^T \hat{m}(z_t) \). Standard errors in parentheses. * \( p < 0.10 \), ** \( p < 0.05 \), *** \( p < 0.01 \).
Table C6: Semiparametric estimates by time of month and season (linear part)

<table>
<thead>
<tr>
<th>Share of dairy sales to Tanykina</th>
<th>Time of the month</th>
<th>Season</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>First and last weeks</td>
<td>Middle two weeks</td>
</tr>
<tr>
<td>Health symptoms in household</td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td></td>
<td>-0.021</td>
<td>-0.038</td>
</tr>
<tr>
<td></td>
<td>(0.024)</td>
<td>(0.026)</td>
</tr>
<tr>
<td>Household has insurance coverage</td>
<td>-0.001</td>
<td>0.021</td>
</tr>
<tr>
<td></td>
<td>(0.037)</td>
<td>(0.026)</td>
</tr>
<tr>
<td>... X Health symptoms in household</td>
<td>0.077*</td>
<td>0.044</td>
</tr>
<tr>
<td></td>
<td>(0.040)</td>
<td>(0.034)</td>
</tr>
<tr>
<td>Log milk production this week</td>
<td>0.103***</td>
<td>0.116***</td>
</tr>
<tr>
<td></td>
<td>(0.026)</td>
<td>(0.027)</td>
</tr>
<tr>
<td>Log food expenditures in 1,000 Sh</td>
<td>-0.007</td>
<td>0.072***</td>
</tr>
<tr>
<td></td>
<td>(0.015)</td>
<td>(0.028)</td>
</tr>
<tr>
<td>Village-week median log milk price in market</td>
<td>-0.166**</td>
<td>-0.025</td>
</tr>
<tr>
<td></td>
<td>(0.083)</td>
<td>(0.041)</td>
</tr>
<tr>
<td>... (lagged)</td>
<td>0.197**</td>
<td>0.024</td>
</tr>
<tr>
<td></td>
<td>(0.083)</td>
<td>(0.033)</td>
</tr>
</tbody>
</table>

Center-month effects

<table>
<thead>
<tr>
<th></th>
<th>Yes</th>
<th>Yes</th>
<th>Yes</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>R-squared within households</td>
<td>0.157</td>
<td>0.163</td>
<td>0.203</td>
<td>0.123</td>
</tr>
<tr>
<td>Mean dependent variable</td>
<td>0.500</td>
<td>0.496</td>
<td>0.506</td>
<td>0.493</td>
</tr>
<tr>
<td>Number of observations</td>
<td>1319</td>
<td>1554</td>
<td>1100</td>
<td>1773</td>
</tr>
<tr>
<td>Number of households</td>
<td>88</td>
<td>88</td>
<td>81</td>
<td>86</td>
</tr>
</tbody>
</table>

Notes: The analysis is restricted to households with variation in the share of milk sold to Tanykina, and weeks in which these households sell milk. The log of variable $x$ is proxied by $\log(x + 1)$. We report $R^2$ in the semiparametric regressions by the squared correlation coefficient between the demeaned $y$ (minus household mean across time) and fitted value of demeaned $y$, $\tilde{y}_i$; $\tilde{y}_i = y_i - \bar{y}_i$ and $\tilde{y}_i = (x_i - \bar{x}_i)\hat{\beta} + \hat{m}(z_i) - \hat{m}_i$, where $\bar{y}_i = \sum_{t=1}^{T} y_{it}$, $\bar{x}_i = \sum_{t=1}^{T} x_{it}$, and $\bar{m}_i = \sum_{t=1}^{T} \hat{m}(z_{it})$. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. 

73