Incomplete Credit Markets and Commodity Marketing Behavior*

Emma C. Stephens†
Assistant Professor
Economics Field Group
Pitzer College

Christopher B. Barrett
International Professor of Agriculture
Dept. Applied Economics
and Management
Cornell University

June 2009

*We acknowledge with gratitude support from the Rockefeller Foundation, the United States Agency for International Development (USAID), through grant LAG-A-00-96-90016-00 to the BASIS CRSP and the Strategies and Analyses for Growth and Access (SAGA) cooperative agreement, number HFM-A-00-01-00132-00, and the Coupled Natural and Human Systems Program of the Biocomplexity Initiative of the National Science Foundation, through grant BCS # 0215890, as well as the Tegemeo Institute of Agricultural Policy and Development, which graciously made the data available. We are also grateful to Marc Bellemare, Diansheng Dong, George Jakubson, Hyejin Ku, Fidan Kurtulus, Carol Murphree, Chuck Nicholson, Kerry Papps, Elisheba Spiller, Steven Yen and audiences at Buffalo State College, Cornell University, University of Delaware, Pitzer College and at the 2006 annual meetings of the American Agricultural Economics Association and the Northeastern Universities Development Consortium for invaluable comments and suggestions. The views expressed here and any remaining errors are the authors' and do not represent any official agency.

†Corresponding author. Pitzer College, 1050 N. Mills Ave., Claremont, CA 91711. Tel: 909-607-7952. Fax: 909-607-7880. E-mail: emma_stephens@pitzer.edu.

Abstract

We develop a simple theoretical model of market participation over multiple seasons in the presence of liquidity constraints and transactions costs to explain the 'sell low, buy high' puzzle in which certain households forego opportunities for intertemporal price arbitrage through storage and are observed to sell output post-harvest at prices lower than observed prices for purchases in the subsequent lean season. We test our model with data from western Kenya using maximum likelihood estimation of a multivariate sample selection model of market participation. Access to off-farm income and credit indeed seem to influence crop sales and purchase behaviors in a manner consistent with the hypothesized patterns.

JEL Codes: O13, O12, Q12, D91

Keywords: Commodity markets, maize, liquidity, seasonality, Kenya.

1. Introduction

Regular, sharp seasonal price fluctuations are a common characteristic of staple grains markets in many developing countries (Sahn, 1989). Yet many farmers appear not to take advantage of the apparent intertemporal arbitrage opportunities created by predictable seasonal price variation in storable commodities. Instead, they often sell their output at low prices post-harvest and buy back identical commodities several months later for prices far higher than they received post-harvest.[*]

Several candidate reasons exist that might explain this ‘sell low, buy high’ puzzle, which is clearly at odds with unconstrained, intertemporal profit-maximizing behavior (on which, see Williams and Wright (1991)). First, impatience could lead to very low storage rates of staple commodities. However, seasonal price increases often so far exceed prevailing local interest rates that this explanation frequently seems implausible. For example, in Kenya in the 2002-2003 crop year the mean annual change in maize prices across three large market centers (Bungoma, Kisumu and Nairobi) was 44%, while the mean bank deposit rate was only 5% (IGAD, Central Bank of Kenya). Given such patterns and continuous household demand for basic grains for survival, it seems implausible that discount rates could be high enough to broadly explain the ‘sell low, buy high’ puzzle.

Second, appropriate storage technologies might not be available, raising intertemporal storage costs to the point that storing output for future sale becomes unprofitable. But even at high inter-seasonal storage loss rates of 10-30%,[†] the preceding argument still applies; one would need an implausibly high discount rate to make ‘sell low, buy high’ an attractive strategy. Moreover, given the predictability of sharp seasonal price increases and the availability of inexpensive grains storage technologies that reliably exhibit annual loss rates of only 1-2% (Barrett, 1997), there would seem to be high returns to investment in better household-level storage technologies that would obviate such explanations. Furthermore, we routinely observe households investing in retaining calves or other not-

[*] See Barrett (2007). However, this is not a universally widespread phenomenon, as Alderman and Shively (1996) show for Ghana.

[†] These losses could be due either to biophysical deterioration or loss of commodities or to claims made on stored grains by family and friends, i.e., implicit social taxation of storage.

yet-productive livestock, in children's education that will pay off only with a long lag, and in other ventures; thus there is plainly a general willingness to invest, just not necessarily in holding grains stocks interseasonally. The low storage quality explanation of 'sell low, buy high' marketing patterns therefore also seems implausible.

A third possible explanation for low storage demand could be longer term concerns about price risk over several growing seasons. However, Saha and Stroud (1994), Barrett and Dorosh (1996) and Park (2006) have all explored the role of grain storage as a price hedge *ex ante* and find analytical and empirical support for storage patterns that also run contrary to the 'sell low, buy high' puzzle; they show that price risk aversion should generate excessive, rather than insufficient, stockholding post-harvest.

An alternative class of explanations little explored empirically to date in the literature is that the 'sell low, buy high' phenomenon represents a 'displaced distortion' (Barrett, 2007) arising due to financial markets failures that people implicitly resolve through seemingly-irrational resource allocation patterns. If people have no other means of addressing temporary liquidity constraints, they might find it optimal to convert non-cash wealth in the form of grains into cash, even knowing that they will need to buy back grain later at a higher price, with the associated losses representing the de facto interest rate on a quasi-loan for several months. In this paper, we develop that line of argument by combining models of consumption smoothing under liquidity constraints (Schechtman and Escudero, 1977; Zeldes, 1989; Deaton, 1991), with the literature on market participation (de Janvry et al., 1991; Goetz, 1992; Key et al., 2000; Bellemare and Barrett, 2006) to develop a model of commodity marketing and storage that can explain 'sell low, buy high' grain marketing behavior. We interact predictable seasonal market price increases with potential fluctuations in the household's shadow valuation of grains that are both consumed and produced by the household and potentially bought and sold in the market. We show that liquidity constraints can lead to greater seasonal variability in shadow prices for agricultural output and that this instability can lead to seemingly perverse marketing behavior, with liquidity constrained households most at risk for failing to seize intertemporal arbitrage opportunities through seasonal grain storage. We then test the hypotheses generated by this model using data on seasonal maize marketing patterns among farmers in western Kenya.

2. Model Development

The following outlines the household's period-by-period consumption and production decisions, comparing the unconstrained case with the liquidity-constrained scenario.

2.1. Basic household model without liquidity constraints or transactions costs

Equation (1) presents a simplified version of the agricultural household's intertemporal maximization problem, defined over a generalized utility function $U(\bullet)$, with household specific tastes (θ) and time preference (δ).[‡] Assume at time t that the household consumes a staple food commodity (c_t) (that is also produced by the household (Q_t)), as well as a composite tradable good (x_t) that can only be purchased in the market. Therefore, in each period, the household consumes c_t and x_t, produces Q_t which it has the ability to store (S_t), and borrows or saves cash ($\pm b_t$), which it must pay back in the following period. Storage depreciates at a rate of ν and savings earn a constant, riskless interest rate, r, which is also the interest rate on borrowing.

The household uses its own family labor (L_{ft}), hired labor (L_{ht}), land and other fixed factors (A) and a vector of other inputs ($\mathbf{z_t}$) to produce output Q_t through production function $F(\bullet)$. The household can also earn off-farm income at wage rate w for labor L_{oft} and pays hired labor the same parametric wage rate. Total labor availability in the household, L^T, constrains labor allocation. The price of x_t is numeraire, while the price of c_t (and Q_t) is given by p_t and the price of each z_t is given by the vector of prices $\mathbf{p^z}$. Full income, Y_t, constrains consumption and represents the value of all agricultural output and net carryover grain stocks and savings and labor income net of other production input costs. The household must choose a consumption time path for c_t and x_t as well as production inputs, storage and cash savings, so as to maximize its expected lifetime utility given uncertain future prices and the budget and labor availability constraints it

[‡]We abstract from the market participation problem for the moment in order to focus on credit constraints and thus treat prices quite generally, temporarily ignoring transactions costs.

faces:

$$\underset{\{\Gamma_t\}^{\infty}{}_{t=0}}{\text{Max}} E_t \sum_{t=0}^{\infty} \left(\frac{1}{1+\delta}\right)^t U(c_t, x_t|\theta) \tag{1}$$

subject to:

$$x_t + b_t + p_t c_t \leq Y_t \tag{2a}$$

$$Q_t = F(L_{ft}, L_{ht}, z_t|A) \tag{2b}$$

$$Y_t = p_t \left[Q_t + (1-\nu)S_{t-1} - S_t\right] + (1+r)b_{t-1} + wL_{oft} - (\mathbf{p^z z_t} + wL_{ht}) \tag{2c}$$

$$L^T \geq L_{ft} + L_{oft} \tag{2d}$$

$$Y_t, L_{ft}, L_{oft}, L_{ht}, c_t, x_t, z_t, S_t \geq 0 \tag{2e}$$

where $\Gamma_t = \{c_t, x_t, L_{ft}, L_{oft}, L_{ht}, z_t, S_t, b_t\}$ and E_t is the expectations operator at time t.

The behavioral result of dynamic optimization problem (1) is a system of output supply and consumption demand equations (3) that are functions of current and expected future prices, as well as current and future realizations of income, w and Y_t (in the case of consumption demand) and fixed factors of production A (in the case of output supply)(Antle and Capalbo, 2001).[§]

$$\begin{aligned}
c_t &= c(p_t, E_t p_{t+1}, p^z, w, Y_t, E_t Y_{t+1}) \\
x_t &= x(p_t, E_t p_{t+1}, p^z, w, Y_t, E_t Y_{t+1}) \\
q_t &= q(p_t, E_t p_{t+1}, p^z|A)
\end{aligned} \tag{3}$$

We assume that current consumption is decreasing in all own prices as well as in the ratio of current to expected future own prices, and increasing in current and future incomes.

Optimal consumption over time of grain and the market good is found by equating the marginal utility of consumption of c_t and x_t within and across time periods, net of

[§] These functions are also conditional on the household's tastes, discount rate, and local prevailing interest rates. This stylized version of household interseasonal behavior does not incorporate the impact of changing food consumption on income, as in Behrman et al. (1997b) discounting (given by $\frac{1+r}{1+\delta}$). For the market commodity x_t, this results in consumption smoothing over time following the familiar Euler equation: $U'_{x_t} = \left(\dfrac{1+r}{1+\delta}\right) E_t U'_{x_{t+1}}$.

For grain consumption, optimal intertemporal smoothing when there is seasonal price variation requires the household to consider the effects of expected grain price variability as well as marginal utility. The household will still optimally consume c_t and x_t so that the expected marginal utility is equated between goods in each time period. In the current period, this implies that $U'_{c_t} = p_t U'_{x_t}$. Looking into the future, however, the household must take into account the relationship between future grain prices, full income and optimal consumption of the market good to determine grain demand across seasons. This implies that:

$$E_t U'_{c_{t+1}} = E_t \left(p_{t+1} U'_{x_{t+1}} \right) \tag{4}$$

or:

$$E_t U'_{c_{t+1}} = cov \left(p_{t+1}, U'_{x_{t+1}} \right) + E_t p_{t+1} E_t U'_{x_{t+1}} \tag{5}$$

The Euler equation for c_t that incorporates these factors is given as:

$$U'_{c_t} = k_t^{-1} \left(\frac{1+r}{1+\delta} \right) \left(E_t U'_{c_{t+1}} - \Omega_t \right) \tag{6}$$

with $k_t = \frac{E_t p_{t+1}}{p_t}$ and $\Omega_t = cov \left(p_{t+1}, U'_{x_{t+1}} \right)$. For price risk averse households and staple grain subsistence farmers, such as those that characterize our sample, Ω_t is expected to be greater than zero (Saha and Stroud, 1994; Barrett, 1996).

It is easy to see that both predictable seasonal price variation and price risk aversion should lead to high latent demand for grain in the harvest season, either for consumption purposes (c_t) or for storage demand (S_t). For ease of exposition, divide the year into two discrete seasons: harvest (H), at which time market prices are typically low, and lean (L), at which time prices are usually high. Given constant (i.e., no interseasonal variation in)

δ and r, seasonal price variation implies

$$k_H \equiv \frac{E_t p_L}{p_H} > 1 > k_L \equiv \frac{E_t p_H}{p_L} \tag{7}$$

Per the demand functions in (3), and momentarily abstracting from seasonal income differences, period H demand should exceed period L due both to expected seasonal price changes and price risk aversion, as these two conditions imply, under standard assumptions on preferences, that $U'_{c_t} \ll E_t U'_{c_{t+1}}$ at harvest time.

If incomes are higher during H than L, as is typically true, then seasonal income variation merely reinforces this effect. Grain demand should be highest at harvest and lowest in the lean season in this unconstrained setting. As a further simplification, if we assume that grain is costlessly storable, exogenous variation in prices and incomes will then drive demand and marketing patterns. Further, Saha and Stroud (1994) and Park (2006) have both shown that such seasonal price movements should also lead to high levels of storage demand at harvest time as well. The sufficient condition for positive storage in our model implies that:

$$\left(\frac{1-\nu}{1+\delta}\right)\left[\Omega_t + E_t p_{t+1} E_t U'_{x_{t+1}}\right] - p_t U'_x > 0 \tag{8}$$

Under price risk aversion and moderate storage losses ν and discounting, this condition should also hold in the harvest period. The 'sell low, buy high' phenomenon should thus not occur under these assumptions.

As discussed in the introduction, a high discount rate δ, high storage loss rates ν or risk neutrality (such that $\Omega_t < 0$ which might hold for larger scale commercial farmers) could in theory explain low storage rates and harvest demand in the unconstrained intertemporal choice model. However, in economies characterized by sharp, predictable seasonal grain price swings and composed of mostly subsistence farmers, the parameter values necessary to make this condition hold commonly seem implausible. Hence our search for an alternative explanation.

2.2. Household model with liquidity constraints and storage

Following Deaton (1991), one can readily adapt the solution to equation (1) augmented by a liquidity constraint such that $b_t \leq \alpha$. Let $Y(\alpha)$ be full income per (2c) when the liquidity constraint binds. The liquidity constraint implies that if full income Y_t falls below some threshold level, then a liquidity constrained household can do no better than consume all of its available resources. The resulting, kinked Euler equation guiding optimal consumption of c_t over time can now be written as:

$$U'_{c_t} = max\left(U'_{c_t}(Y_t(\alpha)), k_t^{-1}\left(\frac{1+r}{1+\delta}\right)\left(E_t U'_{c_{t+1}} - \Omega_t\right)\right) \tag{9}$$

Thus a discontinuity emerges in the Euler equation when the liquidity constraint binds. If the household's current income realization is above some minimum threshold, $Y_t^*(\alpha)$,[¶] that depends on the borrowing ceiling the household faces, then it will still smooth consumption by equating current and future expected marginal utility. If, however, current income falls below the threshold (due for example to seasonal variation in income and a low borrowing ceiling), the household cannot afford as much current consumption as it would like in an unconstrained world. Therefore, the marginal utility of current consumption exceeds the expected marginal utility of future consumption, inducing the household to do everything it can to minimize this gap, including liquidating carryover stocks. The consumption demand function for c_t can thus be written as follows:

$$c_t = \begin{cases} c(p_t, w, Y_t(\alpha)), & Y_t = Y_t(\alpha) \leq Y_t^*(\alpha) \\ c(p_t, E_t p_{t+1}, w, Y_t, E_t Y_{t+1}), & Y_t > Y_t^*(\alpha) \end{cases} \tag{10}$$

The key point to note about the demand function is that when household income falls below the threshold, future prices and future income no longer affect consumption choice because optimal stock-out breaks the linkage. The household consumes all of its available resources, irrespective of expected change in prices over time. Because the liquidity con-

[¶]Since $Y^*(\alpha)$ requires that $b_t = \alpha$, it must be true that $\frac{\partial Y^*}{\partial \alpha} > 0$, i.e., the threshold increases with the borrowing ceiling. The prior section merely presents the case where $\alpha = \infty$ so that no kink emerges in the Euler equation.

straint creates a kink in the Euler equation that disrupts equilibration of marginal utility across periods, it likewise obviates the standard intertemporal arbitrage conditions that guide household behavior. Hence the basic intuition behind our model of the 'sell low, buy high' phenomenon: that liquidity constrained households may optimally sell when prices are low not because they do not recognize predictable seasonal appreciation in the value of storable grain stocks but, rather, because their current income and expenditure needs force liquidation of their entire asset stock, rendering intertemporal arbitrage opportunities moot.

The presence of physical grain storage in the model does not change this fundamental point. When there exist both physical and financial assets, the household can choose the form in which to hold its wealth: either as cash savings or as stored grain (Park, 2006). The optimal asset allocation then depends on the relative returns to different assets, on risk preferences, etc. This does not change the fundamental, qualitative result of the preceding analysis: a binding liquidity constraint makes storage–whether in cash or in kind–less likely. The value of stored grain is proportional to the marginal value of cash over time. As consumption of c_t falls when the liquidity constraint binds, the current marginal utility of c_t increases, making storage so as to increase expected future cash returns less likely.

The credit-constraint induced reduction in grain storage also has implications for the household latent supply function over time. When the liquidity constraint binds, demand for storage decreases and the household stocks out of its carryover grain supply. Given these two factors, this implies that the latent carry-over supply of grain in the household is extinguished in the periods following a binding liquidity constraint. This situation is reevaluated each harvest season. Due to the fact that realizations of income are stochastic, in seasons where harvest is adequate and the household has no extraordinary expenses, then grain storage is positive (conditional upon preferences and storage costs) and the household can maintain grain supply between seasons so as to take advantage of intertemporal arbitrage opportunities. But as soon as realized income falls below the income threshold, it is no longer optimal to hold grain stocks as a buffer against future income draws and the household consumes whatever existing grain stores it has on hand.

There is thus also a 'kinked supply' curve between harvest realizations that is the analog to the liquidity constrained demand curve shown in (10). Starting the latent supply curve (labeled q_t below) iterations in the harvest season (H) (initial time=t), the impact of a binding liquidity constraint at harvest time on household intraharvest supply is shown in (11):

$$(q_H, q_L) = \begin{cases} (Q(p_H|A) + (1-\nu)S_{t-1}, 0), \;\; Y_H \leq Y_t^*(\alpha) \\ \\ (Q(p_H, E_H p_L|A) + (1-\nu)S_{t-1} - S_H, (1-\nu)S_H), \;\; Y_H > Y_t^*(\alpha) \end{cases} \tag{11}$$

The effect of the credit constraint on supply is thus, likewise, to decouple current and future production decisions and eliminate the use of storage to maximize expected intertemporal income and thus utility.

2.3. Transactions costs, market participation choice and seasonality

Nontrivial transactions costs to market participation are as widespread as liquidity constraints in rural areas of developing countries and similarly create discontinuities in observable behaviors. As de Janvry et al. (1991) explain, transactions costs, τ_t, create a price band around the prevailing market price, p_t. For sellers, the net return per unit output sold (p_{st}) is the market price (p_{mt}) minus transactions costs, (e.g. $p_{st} \equiv p_t - \tau_t$), while for buyers, the net cost per unit purchased is just the market price plus transactions costs, or $p_{bt} \equiv p_t + \tau_t$.[‖]

The household makes its market participation decision whether to be a net buyer of grain ($c_t > q_t$), a net seller of grain ($c_t < q_t$), or autarkic in grain ($c_t = q_t$)[**] in each period based on a comparison of the indirect utility it would enjoy under each option (Key et al., 2000). The indirect utilities are evaluated at the price appropriate to each market participation regime, p_{bt} for buyers, p_{st} for sellers, and the nontradable shadow

[‖] We temporarily ignore fixed transactions costs for ease of exposition.

[**] We assume that the storage decision is implicit in the market participation decision. This is justified by the fact that household latent supply, which contributes to the household shadow price determination, is a function of the household's storage decision.

price, p_t^*, for autarky, determined at the point where internal household demand equals supply, such that $p_{st} \leq p_t^* \leq p_{bt}$.

As discussed previously, in the absence of liquidity constraints, demand (c_t) should vary inversely with seasonal changes in market prices, falling as one moves from the low-price harvest season to the high-price lean season. Thus, abstracting for the moment from seasonal variation in incomes, for shadow prices, $p_H^* > p_L^*$ while for market prices, $p_H < p_L$. Assuming constant transactions costs, if it participates in the market at all, the unconstrained household should engage in canonical intertemporal arbitrage, buying during the low-price harvest season, selling during the higher-price pre-harvest season, or both.

Now consider what happens if the household faces a binding liquidity constraint. The liquidity constrained demand function, $c(p_H, w, Y_H(\alpha))$ from equation (10) obviously varies directly with the borrowing limit α that determines the threshold income level at which the constraint binds, $Y^*(\alpha)$. This happens because any increase in α enables the household to reallocate consumption to the present and thereby reduce the gap in expected marginal utility of consumption across periods. As household demand for c_t falls, p_t^* necessarily falls as well. Thus liquidity constraints reduce the household's shadow price for storable grains. If this effect is sufficiently pronounced, it can induce 'sell low' behavior in the immediate post-harvest period.

Figures 1 and 2 illustrate the effect of liquidity constraints on household market participation decisions in the harvest and lean periods, respectively. The figures show latent household demand and supply in shadow price space for a stylized household facing a particular market price, p_t in the harvest (H) and lean (L) seasons. In figure 1, it can be seen that the liquidity constraint unambiguously lowers the household's latent demand in the harvest season, and thus p_H^*, thereby increasing the likelihood that the household becomes a grain seller. For the comparable unconstrained farmer, p_H^* falls within the non-tradable band around the given market price and thus the unconstrained household optimally chooses autarky.[††] But for the liquidity-constrained households whose shadow

[††]It is also possible that for unconstrained households, latent demand is sufficiently high that grain purchases are the optimal marketing regime.

price has fallen, it becomes optimal to sell grain, with no change in market price, transactions costs or the household supply schedule. Low price sales at harvest are a product of the household's binding liquidity constraint.

[*Figure* 1 *about here*]

Figure 2 demonstrates the effect of the binding harvest period liquidity constraint on latent lean season supply, shadow prices and market participation. In the most extreme case shown, the effect of the constraint is to induce a stockout, thereby reducing latent supply (i.e., carryover stocks) to zero and raising lean season shadow prices, p_L^*. If this effect is large enough, then a credit constraint should lead to increased observations of purchases during the lean season when household stocks are low or gone, in spite of the fact that lean season market prices are higher than the harvest period prices at which the household sold grain due to the binding liquidity constraint. Hence the 'sell low, buy high' phenomenon.

[*Figure* 2 *about here*]

To summarize, given the seasonal nature of storable commodity prices, which typically fall at harvest and rise steadily and predictably over the lean season, 'sell low, buy high' behavior clearly turns on either the simple primitives of high discount rates or storage losses, as discussed previously, or liquidity constraints that bind during the harvest period. This is surely not a ubiquitous condition but may occur with some frequency due to some combination of a poor harvest, highly inelastic demand for non-grain items such as school fees, or medical or ceremonial (e.g., funeral, wedding or religious festival) expenses, each of which effectively reduces discretionary income to the point that it falls below $Y^*(\alpha)$. Higher income households would thus be less likely to sell at post-harvest lows and those with extraordinarily high non-discretionary expenses would be most likely.

The 'buy high' phenomenon, in contrast, depends upon reduced latent grain supplies coupled with high lean season prices. This is especially likely if households stop holding grain stores and stock out due to a binding credit constraint. In the case of individual households, the liquidation of household stores to satisfy consumption requirements can lead to extreme spikes in the household shadow price for grain. If this occurs in the lean season, which seems likely due to primary income realization in the harvest period, then it can push credit-constrained, grain-producing households to become grain purchasers, despite high expected prices. Thus the effect of the credit constraint is transferred to market participation behavior throughout the time between harvest realizations and can lead to rational sales at low prices and subsequent purchases of the same commodity at higher prices. The scenario presented above brings into stark relief the quite distinct expected grain marketing choices of liquidity constrained and unconstrained households.

3. Estimation Strategy

The model above yields clear, testable hypotheses. By reducing latent household demand in the harvest period, liquidity constraints should decrease the likelihood of (low price) harvest season purchases and increase the likelihood of (low price) harvest season sales. Further, households that experience harvest period liquidity constraints should also be more likely to undertake (relatively expensive) lean season purchases and less likely to undertake (high price) lean season sales. We can thus empirically explore the liquidity constraints explanation of the 'sell low, buy high' puzzle by testing those hypotheses.

The econometric challenge is that transactions costs create unobserved market participation thresholds, shadow prices are unobservable, market participation behaviors are surely correlated within (i.e., between autarky, buyer and seller status) and across seasons for a given household, and transaction volume decisions are not independent of households' self-selection into the market. One needs to employ an estimation strategy that will address these thorny issues.

Yen's (2005) multivariate sample selection model (MSSM) allows for simultaneous estimation of separate parameters across multiple market participation equations with potentially correlated error structures.[‡‡] In our case, we apply the MSSM to the market entry and quantity decisions for maize grain sales and purchases in both a harvest and a lean season, i.e., to a system of four market entry decision equations and four censored

[‡‡]Goetz (1992) also estimates a multivariate sample selection model of market participation, but considers only one time period.

market quantity equations per household i that can be summarized as follows:

$$\begin{aligned}
&Entry\ Decision\ Vector = K_{sn,i} = \{k_{HP,i}, k_{HS,i}, k_{LP,i}, k_{LS,i}\}\\
&Marketed\ Quantity\ Decisions = Q_{sn,i} = \{q_{HP,i}, q_{HS,i}, q_{LP,i}, q_{LS,i}\}\\
&k_{sn,i} = \{0, 1\}\\
&s = \{harvest\ season(H), lean\ season(L)\},\ n = \{purchase(P), sale(S)\}
\end{aligned} \tag{12}$$

The entry decision is assumed to depend upon the (time-invariant) covariates that influence the shadow price through latent demand or latent supply as well as factors influencing the household specific fixed transactions costs that impact the size of the price band around the market price (Key et al., 2000). The marketed quantity decisions (conditional upon entry) are functions of the factors that influence household latent demand and supply, but are also functions of the (time-varying) market price net of any proportional transactions costs for the particular type of market participation it chooses as optimal. The full specification for the joint entry and quantity equations is therefore:

$$log(q_{sn,i}) = \begin{cases} x'_{sn,i}\beta_{sn} + \nu_{sn,i} & if\ z'_i\alpha_{sn} + u_{sn,i} > 0 \\ 0 & if\ z'_i\alpha_{sn} + u_{sn,i} \leq 0 \end{cases} \tag{13}$$

Both market entry and marketed quantities are random variables. Market entry is observed if the entry equation (shown as $z'_i\alpha_{sn}+u_{sn,i}$ in (13)) is greater than zero. If market entry is observed, then the quantity transacted in the market is given by $x'_{sn,i}\beta_{sn} + v_{sn,i}$. Both $u_{sn,i}$ and $v_{sn,i}$ are assumed to be mean zero, normally distributed random variables. If we let the concatenated vector of the entry and level equation error terms be represented by e, then the full variance-covariance matrix for the specification is an 8x8 matrix, composed of the variance-covariance matrices of the error terms in the market entry and

The households in our sample are conducting market transactions at very small quantities. Therefore, in practice, we are only considering fixed transactions costs in the entry decision in our estimation.

The logarithmic transform of the marketed quantity variable, q_t is used to avoid having to compute the multivariate likelihood function using a truncated normal distribution for strictly positive marketed quantities. This is a common simplification used in other studies of multivariate censored demand systems (Jones, 2000; Yen, 2005).

level equations, both within as well as between these equations:

$$\sum_{8x8} = E(e'e) = \begin{bmatrix} \sum_{u'u} & \sum_{u'v} \\ \sum_{v'u} & \sum_{v'v} \end{bmatrix}, \quad e \equiv [u_{sn}, v_{sn}] \tag{14}$$

The full likelihood function is identical to that of Yen (2005) and is described briefly in the appendix.

The MSSM estimator was initially developed to identify significant covariates of household decisions to purchase consumer goods as well as the quantity purchased in censored demand systems. One identifies the market entry equation by incorporating covariates thought to affect the discrete market participation decision but not the conditional quantity choice. We use the MSSM estimator in a similar way in order to investigate the significance of household cash liquidity constraints on sales and purchases entry and quantity decisions in the harvest and lean seasons, per the theoretical model of section 2. In our case, the identifying variables implied by theory are the fixed transactions costs associated with market entry, which should be irrelevant to the marketed quantity decision conditional on participation.

We face a second identification challenge because our core hypotheses revolve around how liquidity constraints affect market entry and quantity transaction decisions. We operationalize (the absence of) liquidity constraints using the ability to borrow (i.e., access to credit) and access to steady, significant cash flow associated with off-farm income from salaried or skilled employment (or self-employment). Off-farm income is included in our tests for the effect of liquidity since, in the absence of more formal borrowing, households with consistent cash flow from a salary may nonetheless also be able to avoid the 'sell low, buy high' marketing pattern as this income is likely readily available for non-market purchases (for example through holding cash savings). In terms of our model, households with a better ability to cover all consumption expenses adequately, either through readily available cash or borrowing, have incomes consistently above the threshold and therefore avoid such mis-timed marketing of their agricultural output.

Finally, because access to credit is likely endogenous, we need to instrument for it before testing our core hypotheses. We identify the instrumenting regression for credit access using covariates likely to reflect lenders' costs of extending credit to a given household and other transactions cost measures likely associated with credit access but not with maize market participation. We now explain the data and these variables.

4. Data

The data, collected by the Tegemeo Institute of Agricultural Policy and Development, come from a 2005 survey of 1682 households in 4 Districts across 137 villages in western Kenya and report on many aspects of household production, consumption and marketing behavior, including monthly purchases and sales (and associated prices) over the course of the previous year (i.e. from July 2004 to June 2005). The survey data also contain information on local commodity markets as well as market-based interventions like cereal banks, market information initiatives and a program designed to increase agricultural credit by extending credit to agricultural input retailers. So as to increase the observation of rare events (e.g., cereal bank membership), the survey design was choice-based rather than a strictly random sample. Therefore, all statistical analysis has been appropriately reweighted to account for the sampling design using techniques found in Manski and Lerman (1977). Summary statistics for the liquidity and market participation variables appear in table 1. Other key household statistics can be found in the appendix.

[*Table* 1 *about here*]

Most farmers in the sample engage in rainfed agriculture, on farms of three acres or less. Households grow maize, the staple crop, and either sell or store it on the farm until it is either consumed or sold in the period between harvests. Typical storage facilities for maize are open bins constructed of wood or bamboo that are raised off the ground

Further information about the institute can be found on their website: www.tegemeo.org

Households also grow some cash crops. Due to possible endogeneity in the crop portfolio decision and a lack of suitable instruments, we are unable to directly control for these other crops in our specification. However Park (2006) has demonstrated that households growing both grain and cash crops should still primarily use their grain stocks as a hedge against consumption risk. Thus we believe that the presence of cash crops in our sample should not interfere with testing our main hypotheses concerning grain marketing.

to protect the output from pests, or simply in bags inside the family home. In rare cases, a household will have a concrete storage area for grains. Households that belong to cereal banks store some of their output as a share contribution held at the cereal bank, which typically occupies a concrete structure in the local market place. However, the combination of these storage technologies appear to be sufficient to protect households from high storage losses, as over 87% of the maize growing households in the sample report zero losses of harvested maize due to spoilage and the average rate for those experiencing any losses was less than 8%.

4.1. Seasonal market participation

The data collected summarizes monthly marketing patterns for the households. However, to make the estimation more tractable and to limit the number of zero observations, we aggregated household market participation into a single, average harvest period and lean period. Kenya's western region has bimodal rainfall, with a 'long rains' season that runs from April to June (with long rains harvests beginning in July) and a 'short rains' season from October to November (with harvests from November to January). For this analysis, we divided the data into a broadly defined 'harvest period' (running from July to January) and a 'lean period' from February to June, although it technically encompasses two distinct growing seasons. We did this because survey data indicates that over 80% of the households had no stored maize grain at the time of the survey (which occurred at the end of the short rains season) and that most had run out during the month of February. Thus, the stock-out behavior we wish to study did not occur with great frequency in the period between the long and short rains season. We then calculated an average sale and purchase quantity for each period for households that participated and used these averages in the market participation estimation. We also divided the off-farm income into seasonal averages, so that the estimation represents household average behavior for harvest and lean period transactions.

Defining seasons in this manner also increased the total number of observations of each of the four types of market participation.

Lean season prices were higher than harvest period prices and the purchase price — sales price margin was greater in the harvest season as well (table 1 and figure 3); these seasonal differences are statistically significant. Also, according to monthly maize price data available from the Eastern Africa Grain Council's Regional Agricultural Trade Intelligence Network (RATIN, 2009) between 1997 and 2007 for Kisumu, a nearby large market, the average seasonal price spread for maize over these 11 years was approximately 21%, with positive spreads of between 6% and 75% in 9 out of the 11 years. Thus the seasonality evident in our sample does not seem to be particular to the data year.

[*Figure* 3 *about here*]

For our sample, given mean purchase-sales price differences, there was little money to be made by farmers who bought maize at harvest, stored it for a few months, and resold it in the lean season. But those who sold in the harvest season and bought maize back in the lean season faced an average loss of 29.3% (KSh17.393/KSh13.462) on the interseasonal terms of trade, far greater than prevailing local interest rates for those with access to credit. Hence the 'sell low, buy high' puzzle.

Yet many indeed follow that practice. Table 2 summarizes households' net maize marketing position per season. Most households were either net buyers of maize in both the harvest and lean seasons, or net buyers only in the lean season. But these pure net buyers aside, the most common pattern was 'sell low, buy high,' precisely the puzzling pattern we seek to explain. Nearly one in five households was a net seller in the low-price harvest season and a net buyer in the high-price lean season. Ten percent of the sample was autarkic in both periods, neither buying nor selling maize. Other combinations of seasonal purchase and sales behaviors were practiced by less than ten percent of the sample. Only 2% of observations exhibit canonical intertemporal grain price arbitrage, involving purchases in the low-price harvest period and sales in the high-price lean season. Of the nearly 30% of the sample that were net sellers in the harvest period, an astonishing 62% were net buyers a few months later, raising the obvious question of why they would choose

In the two years with negative spreads, prices fell between 4-14%.

a nearly 30% loss on the maize they sold immediately post-harvest and then bought back in the lean season.

[*Table* 2 *about here*]

4.2. Household credit access

Households were asked about credit received (in cash or in kind) for agricultural inputs as well as any credit obtained for non-agricultural purposes. For non-agricultural credit, households were asked whether or not they tried to apply for a loan to cover any non-agricultural expense (like school fees or another similar item), whether or not they were successful in their application and how much they received. Due to the fungibility of credit, we created a single dummy variable indicating reported credit use, whether for agricultural or non-agricultural purposes. We use credit use as a proxy for credit access.

It is important to note that we cannot precisely distinguish the credit constrained from the non-constrained given the data available. More informal sources of credit, such as extended family or local community groups, are not covered by the data and we do not know if people who received credit were nonetheless quantity rationed in the volume received or if those who did not receive credit had no need for it. Our credit use variable is thus an imperfect proxy for the true variable of interest–(the absence of) liquidity constraints by virtue of credit access but is the best available, given the data.

To account for the possible endogeneity of the credit use dummy variable, we predict credit use probabilities for each household in the sample using a probit model instrumenting regression, and then use these predicted values in the second-stage MSSM model of market participation decisions. We follow Kochar (1997) and characterize this first stage equation as representing the difference between a household's marginal return from using credit and the marginal cost to a lender of providing it. We therefore include several household demographic variables and distance measures to control for household credit demand as well as variables that are associated with potential household collateral (in-

Distances from financial institutions have been shown in previous work to affect borrowing behavior (Behrman et al., 1997a) and thus we are using distances to markets, piped water, etc. as proxies for possible distances to other kinds of infrastructure and institutions, like banks.

come and land owned) and monitoring (experience) that might reduce the costs of credit for lenders as identifying variables.

The results of the first stage probit estimation on credit usage are included in the appendix (Table A.2). As one would expect, credit use is strongly and statistically significantly increasing in household income, in household labor endowments, longevity in the village, in the educational attainment of the household head. The identifying vector of distance variables is jointly statistically significant and generally exhibits the expected, negative point estimates, indicating proximity to places where one commonly finds (typically micro) financial institutions in rural western Kenya fosters greater credit access. The one curious exception is distance to a health center, which is positively and significantly correlated with credit use, which likely reflects the low density of health centers in the region.

5. Econometric Results

As indicated previously, the core hypotheses of interest concern the coefficient estimates on our two variables reflecting household liquidity: predicted probability of credit use and household off-farm income. With respect to the four market entry equations, we expect that household liquidity should reduce the probability of harvest period sales and lean season purchases, and increase the probability of harvest season purchases and lean season sales. The literature on market participation indicates parallel predictions with respect to the volume equations, although without detailed income data to fully control for income effects, we are less confident about those point estimates than about the market entry decision estimates. Tables 3-4 display the estimates from the entry and quantity equations, respectively, while tables 5-7 contain the various cross-equation covariance estimates between the entry and quantity equations.

[*Tables* 3 *and* 4 *about here*]

Off-farm income could be endogenous to market participation and volume decisions as well. But since we use exclusively salaried and skilled, year-round employment, this is likely predetermined when households made their 2005 marketing decisions. Moreover, the data have no suitable instruments to identify this variable separately from the market entry and credit use variables, so we have no viable options for resolving any prospective endogeneity in the off-farm income variable.

We obtained qualitatively similar results with a univariate specification for each of the market participation equations, which are available upon request. They have been omitted here in the interets of

The coefficient estimates on our two measures of liquidity, predicted credit access and off-farm income, are jointly consistent with our hypothesis that households with sufficient access to liquidity successfully avoid selling low and buying high in the maize market. Credit use is associated with reducing the likelihood of market entry as sellers in the harvest period and off-farm income is associated with reduced likelihood of purchases in the lean period. Further, predicted credit use is significantly associated with an increased likelihood of harvest season purchases. The lack of significance for either liquidity measure on lean period sales is likely due to the fact that sample households are overwhelmingly net maize buyers and thus sales of any sort are not expected. The different forms of household liquidity have similar estimated effects on marketed quantities. Households using credit and with larger off-farm incomes transact more in the market than those without. This may well pick up omitted income effects. Overall, while the results are not entirely clear cut, the evidence clearly supports the hypothesis that liquidity constraints drive households to practice the 'sell low, buy high' maize marketing strategy in rural Kenya.

Our confidence in the parameter estimates of interest concerning the effects of liquidity on grain marketing is buttressed by the fact that the other parameter point estimates are also consistent with expectations. For example, the entry coefficients for total acres owned show that households with more land holdings are more likely to sell and less likely to buy maize in either season, and once part of the market, these same households tend to transact in larger quantities than those with smaller land sizes with the exception of harvest period purchases. This makes intuitive sense if these households are simply producing more on their land and therefore enjoy larger marketable surpluses. This surplus allows larger sales quantities in both periods, as well as greater ability to consume out of own production at harvest time. For lean period purchases, these households may also benefit from greater income earned during the year, which may boost lean period purchases.

---

brevity. We estimated the multivariate model using GAUSS 9.0. The necessary multivariate cumulative distribution functions were evaluated with the GHK simulator (Hajivassiliou, 1997).

Prices influence marketed quantities in the manner expected except for lean period sales, but again, this is likely due to the fact that lean period sales are not frequently observed. The signs of the estimates on greater storage capacity suggest that households with better storage facilities tend to participate less in the market overall, although this relationship is not statistically significant. This fact aligns with previous work on the role of grain storage already cited above (Barrett and Dorosh, 1996; Park, 2006).

The cross-equation covariances provide additional evidence on the relationships between harvest and lean period market participation decisions. Although the covariances between entry equations are not well identified, the diagonal elements in the matrix of covariances from Entry-to-Level equations are analogous to the inverted Mills ratios that are often calculated as part of typical univariate Heckman models of sample selection. We can therefore see from these estimates that households in the market in general transact more than a randomly selected household, as all of the entry-to-level covariances on the diagonal are positive. We can also see this by looking at the Level-to-Level covariances, which are all positive, indicating that both sales and purchase quantities are positively correlated for market participants. Overall, the statistical significance of many of the cross-equation covariance estimates underscores the importance of estimating these behavioral equations using a systems approach such as MSSM.

[$Tables\ 5-7\ about\ here$]

A few of the parameter estimates run counter to intuition. The mostly positive point estimates on transactions costs seem to indicate that households farther from markets are more likely to make transactions in markets. However, Renkow et al. (2004) found little relationship between distance to market and transactions costs for villages in Kenya without access to motorized transport. It is possible therefore that our parameter estimates are capturing some other features of our sample villages such as those that lie behind the results in Renkow et al, and our distance variable is thus only an imperfect proxy for fixed transactions costs. However, we are not able to include specific transport types in our estimates, and are limited in our ability to further explore this issue.

A final point on the presented estimates concerns the validity of the standard errors. As we have used predicted credit as a regressor in the final estimation of the MSSM, an appropriate solution is typically the use of bootstrapping techniques (Horowitz, 2001) to produce consistent standard errors. However, given the fact that the dependent variable in the first stage is binary, bootstrapping in our case produced a high number of samples that could not be used to estimate the model, as there were no observations of individuals with access to credit in the replication sample produced by the bootstrap procedure. Therefore, the standard errors have not been corrected to account for the presence of the generated regressor, due to the infeasibility of the usual corrective with our particular model, and the absence of other acceptable alternatives. The fact that our univariate results on the main parameters of interest are qualitatively similar to those presented here suggests that this should not greatly alter the inferences made from the multivariate model.

6. Conclusions

This paper empirically explores the oft-observed 'sell low, buy high' puzzle of smallholder food marketing behavior based on the hypothesis that liquidity constraints drive poor households to use commodity markets as a substitute for financial markets to which they have limited or no access. Although considerable, predictable seasonal increases in grain prices should dissuade households from selling staples at low prices post-harvest and buying them back again a few months later, we find that 18% of a recent sample of smallholder households in rural western Kenya in fact practice the 'sell low, buy high' strategy. As noted by Park (2006, pg. 1088), grain stores maintained from harvest to harvest are typically used as a price hedge to ensure adequate consumption. By contrast, a 'sell low, buy high' behavior between harvests would seem to reflect not only an inability to hedge, but liquidity constraints that compel households to quasi-borrow by liquidating physical grain inventories in an interseasonally unprofitable fashion.

Using an adaptation of a recently developed censored demand systems estimator, we reject the hypothesis that liquidity has no effect on household marketing patterns in favor of the alternate hypothesis that it indeed reduces the likelihood of selling low or buying high. While the quantity parameter estimates vary depending on the kind of liquidity to which the household may have access, the market entry parameter estimates that are more reliable in these data are broadly consistent with the model we lay out. Other parameter estimates largely make sense as well.

The practical concern, of course, is that households who engage in 'sell low, buy high' behavior use up scarce resources in costly grain market transactions, making it more difficult for them to accumulate resources necessary to invest in productive assets or improved technologies so as to sustainably increase incomes. Thus not only do these seasonal flow reversals reflect lower welfare just as they do at more aggregate levels (Barrett and Dorosh, 1996, pg. 636) they also reflect displaced financial market failures that can trap households in long-term poverty through distorted grain marketing patterns.

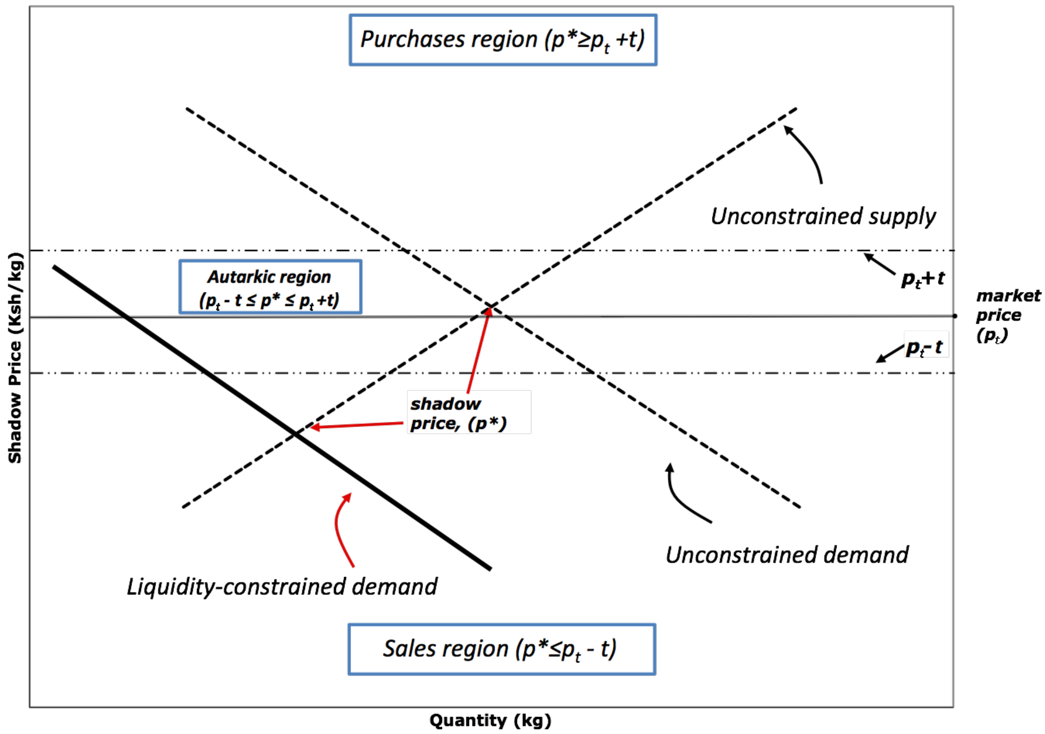

Fig. 1. The impact of liquidity constraints on latent demand and market participation choice in the harvest season ('sell low ')

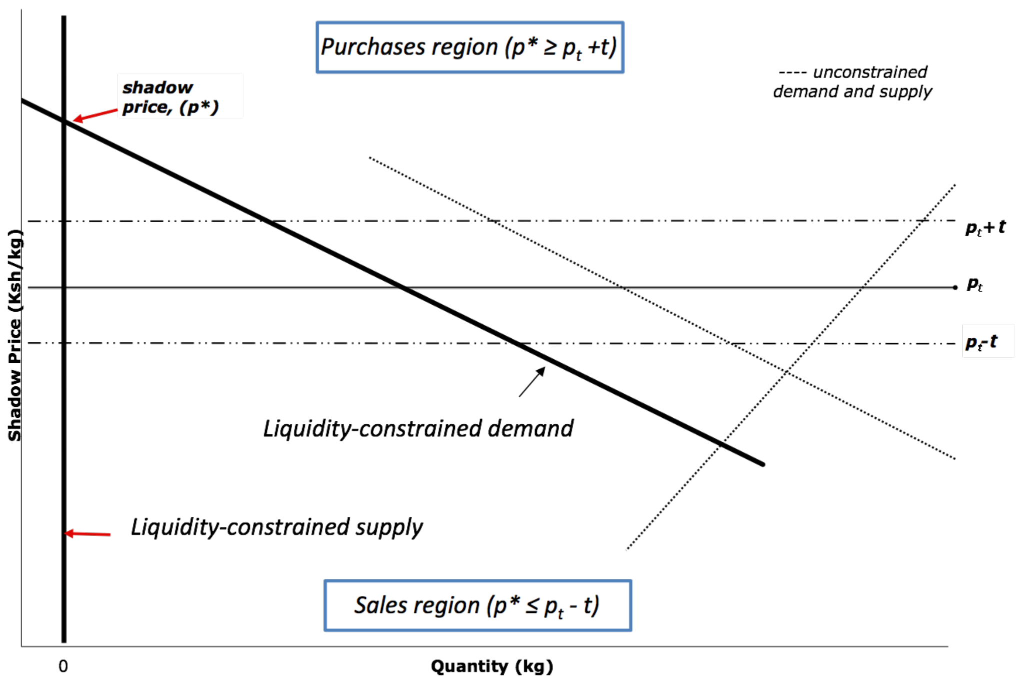

Fig. 2. The impact of liquidity constraints on latent supply and market participation choice in the lean season ('buy high ')

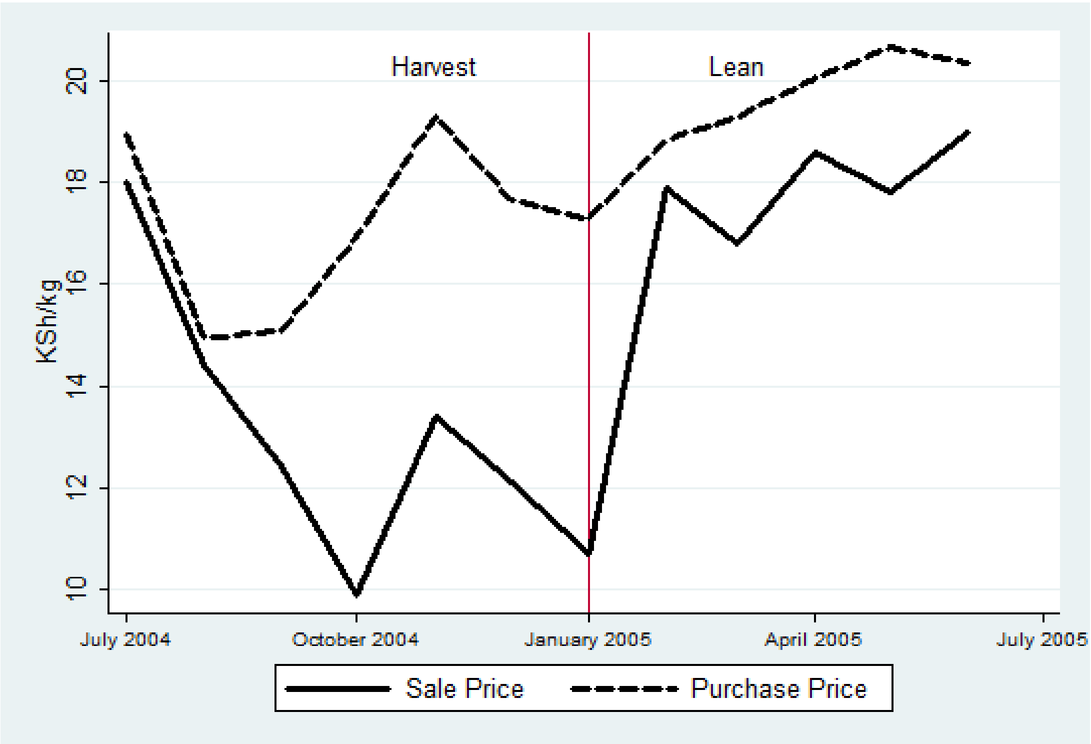

Fig. 3. Average maize sale and purchase prices for the overall sample (N=1682)

Table 1
Summary statistics

Variable Name	**Mean**	**Linearized s.e.**[a]
Predicted Prob. of Non-Agricultural Credit Usage	0.285	(0.004)
Total Off-Farm Income (x100000 KSh)[b]	0.032	(0.002)
Value of grain storage unit (x1000 KSh)	0.859	(0.319)
Distance to nearest shopping center (km)	1.727	(0.035)
Prices[c]		
Harvest Season Maize Grain Purchases (KSh/kg)	16.331	(0.093)
Harvest Season Maize Grain Sales (KSh/kg)	13.462	(0.115)
Lean Season Maize Grain Purchases (KSh/kg)	17.393	(0.046)
Lean Season Maize Grain Sales (KSh/kg)	15.702	(0.179)
Marketed Quantities		
Harvest Season Maize Grain Purchases (kg)	58.130	(2.202)
Harvest Season Maize Grain Sales (kg)	319.227	(22.261)
Lean Season Maize Grain Purchases (kg)	55.892	(1.415)
Lean Season Maize Grain Sales (kg)	598.869	(62.839)

[a]Linearized s.e.s reported to account for sample design on calculation of sample means.
[b]Off-farm income is measured over the entire year for all members of the household who had any kind of off-farm employment
[c]Prices and quantities are averaged only over market participants.

Table 2
Frequency of maize marketing regimes

Marketing Regime (harvest-lean)	**Frequency**	**Percentage of Sample**
Net Buyer-Net Buyer	550	33
Net Seller-Net Buyer	300	18
Autarkic-Net Buyer	327	19
Net Buyer-Net Seller	38	2
Net Seller-Net Seller	73	4
Autarkic-Net Seller	79	5
Net Buyer-Autarkic	36	2
Net Seller-Autarkic	114	7
Autarkic-Autarkic	165	10
	N=1682	*100%*

Table 3
MSSM Estimates of Market Entry by Season

	Entry Decision Equations[a]							
	(HP)		(HS)		(LP)		(LS)	
Liquidity Measures								
Predicted Probability of Credit Use	2.918	(1.200)**	-3.277	(1.194)***	-0.7142	(1.209)	-3.351	(2.075)
Off-Farm Income (1xE6 Ksh)	-8.899	(2.832)***	-2.839	(2.751)	-7.579	(2.654)***	5.308	(3.395)
Supply Shifters								
Total Acres Owned	-11.766	(2.323)***	8.301	(1.882)***	-7.413	(1.864)***	7.279	(2.567)***
Weather Shock (1=planted late)	-0.439	(0.355)	0.410	(0.352)	-0.329	(0.332)	0.948	(0.603)
Storage Container Value (x1000 Ksh)	-2.614	(31.061)	-8.088	(22.509)	-15.753	(26.262)	27.431	(38.962)
Demand Shifters								
Age Household Head (Yrs)	-0.141	(0.765)	-1.480	(0.735)**	-0.691	(0.684)	-1.089	(1.247)
Gender Household Head (1=Male)	-0.082	(0.371)	-0.054	(0.386)	-0.772	(0.364)**	0.851	(0.730)
Dependency Ratio	1.581	(1.224)	0.717	(1.255)	-0.120	(1.273)	0.547	(2.218)
Some formal education (1=yes)	0.268	(0.424)	0.484	(0.458)	0.489	(0.389)	0.199	(0.958)
More than high school (1=yes)	0.154	(0.755)	1.199	(0.760)	0.099	(0.708)	-0.035	(1.382)
Transactions Costs								
Distance to Market Shops (km)	2.497	(0.986)**	0.379	(1.017)	0.948	(0.982)	-0.297	(1.465)

[a]Note: $N = 1682$. Standard errors in parentheses. District dummy variables and the constant included but not reported. Data has been scaled to remain within the range {0,1} as follows: $x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}}$ (Lapedes and Farber, 1988). *,**,*** indicates significant at 10%, 5% and 1% level, respectively. The entry equation variances have been normalized to 8 to allow for numerical evaluation of the likelihood function. Mean log likelihood=-10122.

Table 4
MSSM Estimates of Market Quantity by Season

	Quantity Decision Equations							
	(HP)		(HS)		(LP)		(LS)	
				Liquidity Measures				
Predicted Probability of Credit Use	1.255	(0.196)***	0.143	(0.606)	1.035	(0.160)***	-0.116	(0.755)
Off-Farm Income (1xE6 Ksh)	1.326	(0.568)**	1.192	(1.286)	1.258	(0.382)***	1.731	(1.536)
				Supply Shifters				
Total Acres Owned	-0.047	(0.470)	2.864	(1.195)**	0.327	(0.277)	2.793	(1.230)**
Weather Shock (1=planted late)	0.100	(0.065)	0.041	(0.146)	0.126	(0.051)**	0.321	(0.292)
Storage Container Value (x1000 Ksh)	-1.185	(6.752)	0.601	(8.722)	4.075	(5.148)	-3.365	(3.158)
				Demand Shifters				
Age Household Head (Yrs)	0.203	(0.134)	0.518	(0.338)	0.242	(0.103)**	0.028	(0.491)
Gender Household Head (1=Male)	0.304	(0.061)***	0.339	(0.154)**	0.239	(0.052)***	0.056	(0.339)
Dependency Ratio	0.299	(0.231)	-0.454	(0.480)	0.171	(0.171)	-1.251	(0.851)
Some formal education (1=yes)	0.030	(0.073)	0.615	(0.209)***	0.120	(0.055)**	0.991	(0.337)***
More than high school (1=yes)	-0.331	(0.131)**	1.024	(0.336)***	-0.190	(0.106)*	1.432	(0.502)***
				Market Prices				
Prices (KSh/kg)	-1.163	(0.173)***	1.773	(0.445)***	-1.084	(0.200)***	0.162	(0.697)

Table 5
Estimated Error Covariances: Entry-to-Entry

	Entry-to-Entry $E(u_{sn}, u_{sn'})$ $sn, sn' = \{HP, HS, LP, LS\}$			
	u_{HP}	u_{HS}	u_{LP}	u_{LS}
Harvest Purchase	—-	-0.999	0.892	-0.999
Entry (u_{HP})		(0.928)	(0.896)	(1.532)
Harvest Sales	—-	—-	-0.888	0.994
Entry (u_{HS})			(1.104)	(1.779)
Lean Purchase	—-	—-	—-	-0.999
Entry (u_{LP})				(1.080)
Lean Sales	—-	—-	—-	—-
Entry (u_{LS})				

Table 6
Estimated Error Covariances: Entry-to-Level

	Entry-to-Level $E(u_{sn}, v_{sn'})$ $sn, sn' = \{HP, HS, LP, LS\}$			
	u_{HP}	u_{HS}	u_{LP}	u_{LS}
Harvest Purchase	0.703	-0.586	-0.703	-0.703
Level (v_{HP})	(0.366)*	(0.105)***	(0.103)***	(0.238)***
Harvest Sales	-0.900	1.081	-0.623	-0.113
Level (v_{HS})	(0.186)***	(2.916)	(0.171)***	(0.372)
Lean Purchase	-0.648	-0.374	0.648	-0.648
Level (v_{LP})	(0.092)***	(0.090)***	(0.269)**	(0.138)***
Lean Sales	-1.051	-0.110	-1.051	1.051
Level (v_{LS})	(0.356)***	(0.362)	(0.243)***	(1.806)

Table 7
Estimated Error Covariances: Level-to-Level

	Level-to-Level $E(v_{sn}, v_{sn'})$ $sn, sn' = \{HP, HS, LP, LS\}$			
	v_{HP}	v_{HS}	v_{LP}	v_{LS}
Harvest Purchase	0.495	0.015	0.286	0.094
Level (v_{HP})	(0.030)***	(0.088)	(0.027)***	(0.129)
Harvest Sales	—-	1.168	0.095	0.592
Level (v_{HS})		(0.285)***	(0.059)	(0.113)***
Lean Purchase	—-	—-	0.421	0.052
Level (v_{LP})			(0.018)***	0.092
Lean Sales	—-	—-	—-	1.105
Level (v_{LS})				(0.216)***

Appendix

Table A.1. Summary Statistics

Variable Name	**Mean**	**Linearized s.e.**
Age Household Head (years)	51.380	(0.427)
Gender Household Head (1=Male)	0.798	(0.011)
Dependency Ratio[a]	1.011	(0.023)
Head has some formal education[b] (1=yes)	0.751	(0.012)
Head has more than high school (1=yes)	0.082	(0.007)
Total acres owned	2.310	(0.074)
Weather shock (1=yes)[c]	0.167	(0.010)
Distance to nearest fertilizer seller (km)	3.642	(1.222)
Distance to nearest seller of hybrid maize seed (km)	2.251	(0.049)
Distance to a tarmac road (km)	4.368	(0.119)
Distance to the health center (km)	2.496	(0.052)
Distance to electricity (km)	3.411	(0.110)
Distance to public telephone (km)	2.432	(0.054)
Distance to obtain extension advice (km)	4.211	(0.091)
Distance to piped water (km)	3.667	(0.11)
District Fixed Effects	*Freq.*	*% of Sample*
Number of Households in Bungoma District	591	35.1
Number of Households in Butere-Mumias District	210	12.5
Number of Households in Siaya District	388	23.1
Number of Households in Vihiga District	493	29.3

[a]Dependency ratio is defined as ratio of children less than 15 plus adults over 65 to all other adults in the household.
[b]Comparison case is household heads without any formal education.
[c]This is a zero-one variable indicating whether or not a household performed any crop planting tasks late because of bad weather.

Table A.2. First-Stage Probit for Probability of Credit Use (=1 if hh obtained credit)

Variable Name	**Estimate**[a]	**s.e.**
Total yearly earnings from HH members who earn salaried wages (x1000 Ksh)	0.002	(0.001)***
Total acres owned	0.024	(0.016)
Years household has been farming in the village	0.008	(0.004)***
Household size (number of members)	0.036	(0.013)***
Age of household head (yrs)	-0.007	(0.004)*
Gender of household head (1=male)	-0.092	(0.109)
Head has no formal education (1=yes)	-0.242	(0.129)*
Head has some formal education (1=yes)	0.348	(0.144)**
Distance to a tarmac road (km)	-0.014	(0.012)
Distance to a health center (km)	0.057	(0.021)***
Distance to electricity (km)	0.008	(0.011)
Distance to public telephone (km)	-0.059	(0.030)*
Distance to obtain extension advice (km)	0.007	(0.014)
Distance to piped water (km)	-0.014	(0.014)
Distance to nearest fertilizer seller (km)	-0.002	(0.000)***
Distance to nearest certified maize seller (km)	-0.038	(0.028)

[a]Note: Robust standard errors in parentheses. *,**,*** indicates significant at 10%, 5% and 1% level, respectively. Division fixed effects included but not reported. N=1682, pseudo R^2=0.1124

Yen's (2005) Multivariate Sample Selection Model (MSSM)

Likelihood Function

The MSSM estimator is a multivariate extension of Heckman's sample selection model. We use this estimator to compare the effects of different key covariates on market participation (selection) as well as market quantity transaction decisions (level) for purchases and sales in both the harvest and lean seasons. This yields a four-dimensional sample selection model, with separate equations estimated for Harvest Season Purchases (regime (r)=1), Harvest Season Sales (r=2), Lean Season Purchases (r=3) and Lean Season Sales (r=4). Every household in our sample thus has a unique four-dimensional market participation regime profile, with each participation type making a distinct contribution to the overall likelihood function.

There are three general classes of interest that cover all the different market participation types: Full participation (i.e., observed sales and purchases in each season), Autarky (no market activity observed in any season) and Mixed participation (some type of market activity, but with at least one zero observation out of the four). The likelihood contribution of each is described below.

Class 1: Full Participation

For this class, in both seasons both kinds of market participation are observed. Therefore the selection equation is always greater than zero ($z'\alpha_r + u_r > 0$, $\forall r,\ r = \{1, \ldots, 4\}$). The likelihood contribution for household i in this class is:

$$L_i = g(v_r)(\prod_{r=1}^{4} q_r)\Phi_4(z'\alpha_r + \mu_{u_r|v_r}; \textstyle\sum_{u_r|v_r}) \tag{15}$$

where $v_r = log(q_r) - x_r'\beta_r$ as per the quantity equation in (3), $g(v_r)$ is the four dimensional marginal multivariate normal pdf of the quantity equations over each season and participation type, r, and $\Phi_4(z'\alpha + \mu_{u|v}; \sum_{u|v})$ is the four dimensional multivariate normal cdf of the entry equation and level equation conditional error terms $(u_r|v_r)$. The conditional mean of this variable is given by $\mu_{u|v} = \sum_{uv} \sum_{vv}^{-1} v_r$. The limits to the integrals in the cdf are the $z'\alpha_r$ plus the conditional mean of the entry equation error terms with respect to the error terms from the level equations, ($\mu_{u|v}$). The cdf also has a conditional variance matrix $\sum_{u|v}$.

Class 2: Zero participation

In the case of the complete autarky, the contribution to the likelihood function is:

$$L_i = \Phi_4(-z'\alpha; \sum\nolimits_{uu}) \tag{16}$$

The limits to the integrals are now $-z'\alpha_r$ and the variance matrix is simply over the entry equation error terms, u_r (shown as $\sum_{uu}$).

Class 3: Mixed participation

If the household participates in only a subset m of the total r regimes, the contribution to likelihood is given by:

$$L_i = (\prod_{j=1}^{m} q_j)g(\hat{v_j})\Phi_4(D(z'\alpha + \mu_{u|\hat{v}}); D'\sum\nolimits_{u|\hat{v}} D) \tag{17}$$

Since only certain marketing regimes have positive quantities transacted in the market, the pdf of the level equations is only evaluated for those periods with positive entry, indicated by the subscript j. Therefore, $\hat{v_j}$ contains the level error terms for those periods with observed market participation, and $u|\hat{v}$ is the conditional random variable for the entry equation error terms conditional upon the level error terms for those periods with entry. D is a 4x4 matrix of signs that determine the sign of the limits of the integration and signs within the covariance matrix and is necessary to evaluate the multivariate cdf. Note that matrix D changes with each market regime profile and accounts for the periods with zero entry. Formally,

$$D = \begin{bmatrix} 2d_1 - 1 & 0 & 0 & 0 \\ 0 & 2d_2 - 1 & 0 & 0 \\ 0 & 0 & 2d_3 - 1 & 0 \\ 0 & 0 & 0 & 2d_4 - 1 \end{bmatrix}, \qquad d_t = \begin{cases} +1, & z'\alpha_t + u_t > 0 \\ -1, & otherwise \end{cases}$$

References

Alderman, H., Shively, G., 1996. Economic reform and food prices: Evidence from markets in Ghana. World Development 24(3), 521–534.

Antle, J. M., Capalbo, S. M., 2001. Econometric-process models for integrated assessment of agricultural production systems. American Journal of Agricultural Economics 83(2), 389–401.

Barrett, C. B., 1996. On price risk and the inverse farm-size-productivity relationship. Journal of Development Economics 51(2), 193–215.

Barrett, C. B., 1997. Food marketing liberalization and trader entry: Evidence from Madagascar. World Development 25(5), 763–777.

Barrett, C. B., 2007. Displaced distortions: Financial market failures and seemingly inefficient resource allocation in low-income rural communities. In: Bulte, E., Ruben, R. (Eds.), Development economics between markets and institutions: Incentives for growth, food security and sustainable use of the environment, Wageningen: Wageningen Academic Publishers.

Barrett, C. B., Dorosh, P., 1996. Farmers welfare and changing food prices: Non-parametric evidence from rice in Madagascar. American Journal of Agricultural Economics 78(3), 656–669.

Behrman, J., Foster, A., Rosenzweig, M. R., 1997a. Dynamic savings decisions in agricultural environments with incomplete markets. Journal of Business and Economic Statistics 15(2), 282–292.

Behrman, J., Foster, A., Rosenzweig, M. R., 1997b. The dynamics of agricultural production and the calorie-income relationship: Evidence from Pakistan. Journal of Econometrics 77(1), 187–207.

Bellemare, M., Barrett, C. B., 2006. An ordered tobit model of market participation: Evidence from Kenya and Ethiopia. American Journal of Agricultural Economics 88(2), 324–337.

Central Bank of Kenya, 2004. Annual Report.

de Janvry, A., Fafchamps, M., Sadoulet, E., 1991. Peasant household behavior with missing markets: Some paradoxes explained. Economic Journal 101(409), 1400–1417.

Deaton, A., 1991. Saving and liquidity constraints. Econometrica 59(5), 1221–1248.

Goetz, S. J., 1992. A selectivity model of household food marketing behavior in Sub-Saharan Africa. American Journal of Agricultural Economics 74(2), 444–452.

Hajivassiliou, V. A., 1997. A library of GAUSS routines for simulating multivariate normal rectangle probabilities and their derivatives (http://econ.lse.ac.uk/staff/vassilis/pub/simulation/gauss/simprog.gcf).

Horowitz, J. L., 2001. The bootstrap. In: Handbook of Econometrics, volume 5, pp. 3159–3225, Amsterdam: Elsevier Science.

IGAD, 2007. Intergovernmental authority on development, market information system database (http://www/adwebtec.it/igadmis/mis/nmis/kenya/timeseries1.html).

Jones, A. M., 2000. Health econometrics. In: Culyer, A. J., Newhouse, J. P. (Eds.), Handbook of Health Economics, volume 1A, chapter 6, pp. 265–344, Amsterdam: North Holland.

Key, N. E., Sadoulet, E., de Janvry, A., 2000. Transactions costs and agricultural household supply response. American Journal of Agricultural Economics 82(2), 245–259.

Kochar, A., 1997. An empirical investigation of rationing constraints in rural credit markets in India. Journal of Development Economics 53(2), 339–371.

Lapedes, A., Farber, R., 1988. How neural nets work. In: Anderson, D. Z. (Ed.), Neural Information Processing Systems, pp. 442–456, New York: American Institute of Physics.

Manski, C. F., Lerman, S. R., 1977. The estimation of choice probabilities from choice-based samples. Econometrica 45(8), 1977–1988.

Park, A., 2006. Risk and household grain management in developing countries. The Economic Journal 116(514), 1088–1115.

RATIN, 2009. Regional agricultural trade intelligence network, monthly price data (http://www/ratin.net/priceinfo.asp).

Renkow, M., Hallstrom, D., Karanja, D., 2004. Rural infrastructure, transactions costs and market participation in kenya. Journal of Development Economics 73(1), 349–367.

Saha, A., Stroud, J., 1994. A household model of on-farm storage under price risk. American Journal of Agricultural Economics 76(3), 522–534.

Sahn, D. E. (Ed.), 1989. Seasonal Variability in Third World Agriculture. Baltimore: Johns Hopkins University Press.

Schechtman, J., Escudero, V., 1977. Some results on an "income fluctuation problem". Journal of Economic Theory 16(2), 151–166.

Williams, J. C., Wright, B. D., 1991. Storage and Commodity Markets. NY: Cambridge University Press.

Yen, S. T., 2005. A multivariate sample-selection model: Estimating cigarette and alcohol demands with zero observations. American Journal of Agricultural Economics 87(2), 453–466.

Zeldes, S., 1989. Consumption and liquidity constraints: An empirical investigation. The Journal of Political Economy 97(2), 305–346.