# Do medium-scale farms improve market access conditions for Zambian smallholders?

William J. Burke, T.S. Jayne and N.J. Sitko[1]

Original submitted October 2018, Revision received April 2019, Accepted July 2019

**Note: This is the peer reviewed version of the following article:**

**Burke, W.J., T.S. Jayne and N.J. Sitko. 2019. Do medium-scale farms improve market access conditions for Zambian smallholders? *The Journal of Agricultural Economics.***

**which has been published in final form at https://doi.org/10.1111/1477-9552.12360. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.**

Abstract:

*This study is motivated by the need to understand how the rise of medium-scale farms in Africa is affecting small-scale farm households. Survey evidence over the past decade has shown a dramatic rise in the prevalence of "medium" sized farms between 5 and 100 hectares, but smaller farms still constitute the vast majority of farms and rural households. Prior evidence highlights a co-evolution between the concentration of landholdings, surplus production volumes, and entry of larger traders into the market. Whether these changes tend to impact on smallholders negatively, crowding them out of markets, for example, or positively, potentially opening new marketing channels, is an empirical question. Using a multi-stage model for maize market participation in Zambia, we find that in areas where medium-scale farms are growing, even the smaller farms are becoming more likely to sell maize, more likely to sell to the private sector, more likely to sell to larger traders, and expected sales amongst sellers are higher. On balance, the growth of medium-scale farms and large-scale traders seems to have positive spillover effects on nearby smallholder marketing options.*

Keywords: smallholder marketing; medium-scale farm growth; large-scale trading; agricultural transformation; sub-Saharan Africa; Zambia; market development.

JEL Classifications: Q13; C34; O13; O17; Q17; Q18

[1] William Burke (contact author: william@afpconsulting-burke.com) is with Agricultural and Food Policy Consulting, Baltimore, MD, USA. T.S. Jayne is University Foundation Professor in the Department of Agricultural, Food and Resource Economics, Michigan State University, USA. N.J. Sitko is with the Agricultural Development Economics Division, FAO, Rome. We are grateful to the participants of a session of the International Association of Agricultural Economists, John Mellor, Talip Kilic, the editor David Harvey, and two anonymous reviewers for comments on earlier drafts. Any errors or omissions are our own.

## 1. Introduction

African agriculture is in a state of rapid transformation. While the rural population is still comprised primarily of small-scale farmers, there is mounting evidence from across the continent of a rapid increase in the number of "medium-scale" or "emergent" farmers.[2] Conservative estimates based on available data show that the share of land under control of these emergent farmers now ranges from 20 percent in Kenya, to 32 percent in Ghana, 37 percent in Tanzania, and as high as 53% in Zambia (Jayne *et al*., 2016).

Despite much conjecture, there is very limited evidence on the ways in which the rise of these medium-scale African farmers affects smallholder farmers. Most of the speculation has centered around smallholders' restricted access to land and potential marginalization resulting from large-scale farm land acquisitions (Anseeuw et al., 2016; Messerli et al., 2014; Nolte & Ostermeier, 2017; Nolte & Sipangule, 2017). These are important concerns, to be sure. However, the emergence of larger farms could also generate benefits for small-scale farms (SSFs). One avenue by which this may occur is through changes in the structure and performance of commodity markets that improve nearby smallholders' access to buyers and services. Sitko, Burke, and Jayne (2018) show that in Kenya and Zambia, investment in commodity value chains by larger-scale grain marketing firms tends to be concentrated in areas where medium-scale farms (MSFs) are becoming more prominent.

The data we will describe in this study are consistent with this story. Amongst farmers selling to the private sector, just 14% of the SSFs (cultivating less than 5 hectares) sell to a larger-scale grain marketing firm. Predominantly large buyers (LBs) are large-scale traders (LSTs), but LBs also include millers, who rely on locally available surpluses to process grain at scale.[3] Thirty-five percent of farms cultivating between 5 and 10 hectares sell to an LB, as do 61% of farms cultivating between 10 and 20 ha. Sitko, Burke and Jayne (2018) suggest medium scale farm and large-scale trader sectors are "co-evolving", each attracting and facilitating the growth of the other. Anecdotal evidence from trader interviews and regression analysis suggests the stronger causal relationship is from MSFs attracting large buyers, who enter markets because of the concentration of larger farms and the surpluses they generate. At the district level, if we regress the share of maize sales to an LB on the lagged MSF share of cultivated land[4], the estimated effect is positive, meaningful, and statistically significant at the 1% level.[5] .

---

[2] Emergent or medium scale farms in Africa are usually defined as having between 5 and 100 hectares (ha) of farmland

[3] Most of our discussion revolves around traders, who are the dominant actors in the private sector. To be clear, though, "large buyers" (or LBs) are large-scale traders and millers, whereas "small buyers" (or SBs) are small-scale traders, other households, or other small and otherwise unclassified transactions. For both large and small buyer groups, though, traders dominate the market: 2/3 of maize sold to SBs is sold to SSTs, and 2/3 of maize sold to LBs is sold to LSTs.

[4] The Zambian government defines farms cultivating less than 5 ha as "small", those farming between 5 and 20 ha as "medium" and those cultivating more than 20 ha as "large". Farms cultivating more than 20 ha are not included in the small & medium scale farm surveys used for this study.

[5] Specifically, the districts where there was no land under cultivation on MSFs in the recent past, the expected share of maize sales to LSTs is effectively nil. On the other hand, at the mean value of MSF share of cultivated land (15.5%), the expected share of maize sales to LSTs in the following years is 9%

The volume of grain traded by LSTs is substantial. Sitko, Burke and Jayne (2018) estimate LSTs purchased 518,000 MT of maize and 917,000 MT of total grains in 2015/16, 37 percent of forecasted maize sales in Zambia that year. This grain was purchased through a variety of channels: roughly 40 percent from medium-scale farmers, 34% from small-scale farmers, 21 percent from other traders, mostly small-scale, while only 3.5% from commercial farmers. Of the total grain purchased, LST respondents estimated that 41% was sold to domestic large-scale milling companies, while 13% was sold for export. The remainder was sold through alternative domestic channels, including animal feed processors, oil crushers, NGOs, and other trading firms.

As these figures suggest, while LSTs may be attracted by the rise of MSFs, the change in the grain marketing channel could also present new opportunities and increased competition for those buying from smaller farmers. On the other hand, it may be that LSTs exploit monopsony power to decrease opportunities for farmers. Ultimately, whether the rise of LSTs and MSFs has positive spillover effects on SSFs is an empirical question.

This study investigates the potential relationship between the rise of medium-scale farms and small-scale farm commercialization using household survey data from Zambia. We answer four specific research questions: (i) are SSFs more likely to sell in areas with a relatively high concentration of land under MSFs; (ii) given that they sell, are they similarly more likely to sell to the private sector; (iii) given that they sell to the private sector, are they similarly more likely to sell to large buyers, which are mainly LSTs; and finally, (iv) how do the expected sales to the private sector change in areas where farmland is concentrated, treating large and small buyers as distinct marketing channels?

We address these research questions using a two-wave panel of nationally representative farm survey data from Zambia collected in 2012 and 2015. We focus on smallholder farmers' sales of maize, which is the predominant agricultural product in Eastern and Southern Africa. We explicitly test the hypothesis that there are spillover effects for smaller farms.

## 2. Data

Household survey data in Zambia come from the Rural Agricultural Livelihoods Survey (RALS) carried out by the Indaba Agricultural Policy Research Institute (IAPRI) in partnership with the Central Statistical Office (CSO) and the Ministry of Agriculture and Livestock (MAL). This is a nationally representative longitudinal survey of smallholder households in Zambia carried out in 2012 and 2015. In total, the pooled sample includes 8,838 households in 2012 and 7,933 households in 2015. The sample covers all districts in Zambia and 442 "standard enumeration areas" defined by the CSO census. Consistent with the CSO standard sampling protocol for smallholder surveys in Zambia, larger farms are systemically over-sampled, but representativeness is maintained by applying inverse probability sampling weights. Full details on the sampling and weighting framework are available from the Government of Zambia (GRZ, 2012) and Megill (2004).

To enable respondents to accurately distinguish between different types of maize buyers, enumerators were trained to identify LSTs according to three defining characteristics when respondents indicated that they sold grain to a trader. First, does the trader purchase greater

volumes of grain than the average trader in the area? Second, does the trader personally come to villages to buy grain or does he/she operate buying points and hire agents to buy on their behalf? Typically, LSTs do not themselves travel to villages to acquire grain. Instead, they buy directly from farmers at buying points in towns, hire agents to travel to the villages to aggregate surpluses, or buy from smaller, "satellite" traders without a formal affiliation. Third, does the trader have a company name or is the trader buying grain as an individual? This question allows us to distinguish between designated agents buying on behalf of an LST and traders buying on their own account, where possible. Where the responses to all three questions consistently indicate that an LST was the buyer, then the buyer type is coded as an LST. Table 1 shows the distribution of our farms by size, trading activity and marketing channel. Note, the alternative to selling to the private sector is selling to the government.

**TABLE 1 ABOUT HERE**

Reviewers of this manuscript raised reasonable concerns regarding the reliability of our trader categorization data, since we rely on farmers and not the traders themselves. Indeed, it is an important caveat to acknowledge this as an area where future studies may offer improvement. That said, it is worth emphasizing that we do believe our data are reasonably accurate and that working together, the farmers and trained enumerators were capable of properly identifying traders by categories as we've described them for several reasons.

First, many LSTs are obviously large-scale. Sitko, Burke and Jayne (2018) provide detailed insights into the attributes of LSTs in Zambia using structured survey data collected from all 24 members of the Grain Traders Association of Zambia that are categorized as "large-scale" in the membership roster. They show that LSTs in Zambia are comprised of both domestic and multinational firms. In total 29% of LSTs in Zambia are multinational, including global firms such as Cargill, as well as African based multinationals such as Export Trading Group, AFGRI, and NWK Agri-services. Otherwise, many of the LSTs in Zambia are relatively new market entrants; the mean and median year that LSTs began buying grain in Zambia is 2008 (Sitko, Burke and Jayne 2018). This coincides with the global food price spike that helped trigger renewed global interest in African agriculture (Thurow 2010).

Second, regarding accurate identification, it is worth noting the survey team included a network of more than 20 supervisors, (professors, late-stage graduate students, and professional survey supervisors) in the field, The supervisors were in daily contact with the enumeration teams and each other, as well as 2 Lusaka-based senior-level researchers. If ever an enumerator felt unsure of their classification (on this or any other issue), a supervisor was contacted to assist with correctly recording data, communicating if necessary with others, including the Lusaka based team. Finally, our results are very robust to the possible measurement errors that may have occurred with respect to categorizing traders. The robustness of our results is discussed at length in an on-line appendix.

For the purposes of this study, and consistent with the CSO designations, we define small-scale farms as those operating up to 5 hectares of land, whereas medium-scale farms are those operating 5 to 20 hectares. Small-scale farms account for 95% of the total sample (Table 1). Because the datasets only contain farms cultivating less than 20 hectares, we confine our analysis

to the potential spillover effects resulting from the presence of medium-scale farms, though it is possible that important spillover effects may also derive from large-scale farms.

**TABLE 2 ABOUT HERE**

Table 2 presents some descriptive statistics for the data used in this study. In addition, Figures B1 and B2 (in the on-line appendix) are maps of Zambia to show the distribution and dynamics of farmland consolidation by district. Figure B1 shows the share of land that was under MSFs (5-20 ha) in 2012, and Figure B2 shows the percentage point changes in the share of land under MSFs between 2012 and 2015. Together, these figures show a substantial amount of variation over space and time that is not obvious in Table 2.

## 3. Conceptual framework

The rise of LSTs and MSFs in Africa since 2000 has been well established (Jayne *et al*., 2016; Sitko, Burke & Jayne, 2018), and suggests that the dominant causal pathway of this coevolution has been the traders' attraction to areas with more MSFs and available surpluses. The hypothesis is that scale economies in grain markets are an important determinant of the location of LST investments. From the standpoint of an LST, the costs of purchasing grain per transaction and per ton purchased are lowest in areas where farmers have large surpluses to sell. By contrast, areas dominated by small-scale farms have an atomistic, geographically dispersed production of small surpluses, which is associated with relatively high costs of procurement. Surplus production in such areas are at levels more suitable to the economies of scale of smaller traders. That said, the rise of LSTs can also attract smaller traders who often act as satellites to larger traders, aggregating purchases and re-selling them further along the value chain.

If this hypothesis holds, as the regression analysis[6] we described in the introduction suggests, it can be further hypothesized that smallholder farms are likely to benefit from a greater local presence of LSTs. Small-scale farmers might enjoy higher prices resulting from increased competition and improved market access resulting from LST investment, and the more efficient cost structure that LSTs may enjoy compared to independent small-scale traders resulting from economies of scale in distribution and access to financing for large purchases.

There are some countervailing reasons to suspect that a strong presence of LSTs in an area may not benefit small-scale farmers. First is the worry that LSTs may be in a position to "crowd-out" small-scale competitors and exert monopsony power over local farmers. Through their access to low-cost financing, risk management instruments, and in some cases vertical integration, LSTs may induce market exit by independent small traders and a decline in competition. Second, LSTs may have quality requirements in terms of color, moisture content, and kernel damage, which in some cases require technology and management practices to which small-scale farmers may not be able to adhere (Fafchamps 2001). Ultimately, the impact of farmland concentration and LSTs on maize sales behavior by smallholder farmers is an empirical question.

[6] Recall this is a regression of the share of sales to LSTs on the lagged share of small and medium scale farmland cultivated on farms larger than 5 ha, where the statistically significant coefficient is 0.323. Notably, this OLS regression result is consistent with alternative model specifications such as using contemporaneous farmland concentration as a regressor or a fractional probit models.

It is worth noting that government crop marketing boards have continued to purchase grain in parts of Africa, including Zambia. Longstanding government demand and price support has encouraged individual investment in medium-scale farms over the years, which has in turn encouraged the generation of large surplus-producing farms (Jayne et al., 2016). While government price policies undoubtedly affect farmer decisions to sell to government or private buyers, they have no obvious influence on farmers' decisions to sell to different types of private sector traders. Moreover, Sitko, Burke and Jayne (2018) show that the relative role of government buying has diminished in recent years. This includes the period covered by the present study, but the decline of the government's role has been even more prominent since (Daily Nation Newspaper 2018).

## 4. Estimation approach

### *4.1 A multi-stage-model for market participation*

To address the four research questions, we estimate a multi-stage model of small (less than 5 ha) farm maize market participation where the key variables are district-level shares of cultivated land that are on farms between 5 and 10 ha (which we call "B-farms" using the parlance of Zambia's government) and those between 10 and 20 ha ("C-farms"). In the model's first stage we estimate the relationships affecting the likelihood of selling maize using the sample of 15,087 households cultivating less than 5ha. Then, amongst the 7,215 selling households, we estimate how factors affect the conditional (on being a seller) likelihood of selling to the private sector (rather than the government) in the second stage. The private sector transactions include sales to SSTs and LSTs as well as millers, other households and relatively few otherwise uncategorized sales. The third stage is to estimate, conditional on being a seller and selling to the private sector, how factors affect the likelihood of selling to a "large buyer" (LB) amongst the 3,285 observations that sell to the private sector. The majority (67.2%) of transactions with LBs are sales to LSTs, accounting for 66.9% of maize sold via this channel. Private sector transactions that are not with an LB are with what we will call "small buyers" (SBs), which includes SSTs (accounting for roughly 69% of transactions and 78% of the maize sold via this channel) and other private transactions. Finally, the fourth stage of the model estimates separately the effects of farm concentration on the expected value of sales to SBs or LBs amongst the farmers who sell to each buyer group using a lognormal specification.

**FIGURE 1 ABOUT HERE**

This entire multi-stage model is illustrated in Figure 1, and although the maximum likelihood estimates of all parameters can be obtained by separate probit and lognormal regressions, it is worth noting the full likelihood function[7], which is:

[7] There are 48 "A" farms that sold to both LB and SB groups. Strictly speaking, the hurdle framework would define these as: sold maize=yes; sold to private sector=yes; sold to LBs=yes; and the fourth stage would include these sales (counting only the quantity sold to LBs). It is worth noting that including these 48 observations in the non-LB quantity regression (counting only the quantity sold to SBs) would not meaningfully affect results.

$$f(w_1, w_2, w_3, y, y_2 | x, \theta) = [1 - \Phi(x_1\alpha)]^{1[w_1=0]} \left[ \Phi(x_1\alpha) \left\{ [1 - \Phi(x_2\beta)]^{1[w_2=0]} \left[ \Phi(x_2\beta) \left\{ \begin{array}{c} \left[ \left(1 - \Phi(x_3\gamma)\right) \frac{\phi[\{ln(y_{SB}) - x_4\delta_{SB}\}/\sigma_{SB}]}{y_{SB}\sigma_{SB}} \right]^{1[w_3=0]} \\ \left[ \Phi(x_3\gamma) \frac{\phi[\{ln(y_{LB}) - x_5\delta_{LB}\}/\sigma_{LB}]}{y_{LB}\sigma_{LB}} \right]^{1[w_3=1]} \end{array} \right\} \right]^{1[w_2=1]} \right\} \right]^{1[w_1=1]} \quad (1)$$

Where $w_1$=1 if the household sells maize and is zero otherwise, $w_2$=1 if the household sells maize to the private sector and is zero if the household sells maize to the government, $w_3$=1 if the household sells maize to an LST or miller, and is zero if the household sells in some other private transaction, $y_{LB}$ is the quantity sold to an LST or miller, and $y_{SB}$ is the quantity sold in other private transactions with small buyers. The variables $x_1, x_2, x_3, x_4$ and $x_5$ are the determinants of $w_1, w_2, w_3, y_{SB}$, and $y_{LB}$ respectively, and $\theta = (\alpha, \beta, \gamma, \delta_{SB}, \sigma_{SB}, \delta_{LB}, \sigma_{LB})$ are parameters to be estimated.

Within each stage, the "conditional" interpretation of estimated relationships is the same as if we were looking at isolated models.[8] For example, the second stage results can be used to compute average marginal effects, which can be interpreted by themselves as the marginal effect on the likelihood of selling to the private sector, "conditional" on the observation being a maize seller. Because these relationships are considered conditional on being part of the sample, or relevant sub-sample, we need not be concerned with selection bias in the statistical sense (Burke, Myers & Jayne, 2015).

The key benefit of a multi-stage model is that we can also estimate "unconditional" relationships. For example, the "unconditional" probability that a farmer sells to the private sector is the product of the expected probabilities predicted in stages 1 and 2 of the model depicted in Figure 1 (and equation 1). We can obtain the marginal effect of a variable on that "unconditional" probability using the chain rule and compute it for any observation, regardless of whether or not they actually are maize sellers. These predictions can be interpreted as the marginal effect "per

[8] We put "conditional" and "unconditional" in quotations because their appropriateness is highly contextual. For example, all predictions are conditional on the right-hand-side variables, even when they are what we refer to here as "unconditional" predictions because they are not conditional on any state of being for the observation. Also, within a 4-stage model framework like this, it's quite possible to generate predictions that are conditional on some state variables, but not others. For example, we could predict the probability of selling to a large buyer conditional on being a maize seller, but not conditional on selling to the private sector, which would be the product of predicted values in stages 2 and 3. It's not immediately obvious what we should call such a prediction.

household" at the population level (for the population that is represented in the first stage, that is, farms smaller than 5 ha in this case).

Incorporating the full multi-stage model, we can predict the "unconditional" expected sales to either an LB or SB for every household, whether they sell any maize or not, as the product of expectations from all four stages. The average of this prediction can be roughly interpreted as the average sales to one of these traders at the population level. The average marginal effect of an explanatory variable on the unconditional sales expectation can thus be interpreted as the expected per capita change in sales at the population level. All of these interpretations are analogous to those described in Burke, Myers & Jayne (2015), though here we're discussing a 4-stage rather than a 3-stage model.

It would be cumbersome to list all of the equations for the conditional and unconditional predictions and the marginal effects we expect variables to have on those predictions. As a representative example, equation 2 shows how we would compute the marginal effect of some variable $x_j$ (with corresponding coefficients that have the *j* subscript) on the "unconditional" expected value of $y_{LB}$:

$$\begin{aligned}\partial E(y_{LB})/\partial x_j = {} & \alpha_j \cdot \phi(x_1\alpha) \cdot \Phi(x_2\beta) \cdot \Phi(x_3\gamma) \cdot \frac{\phi[\{ln(y_{LB})-x_5\delta_{LB}\}/\sigma_{LB}]}{y_{LB}\sigma_{LB}} \\ & +\beta_j \cdot \Phi(x_1\alpha) \cdot \phi(x_2\beta) \cdot \Phi(x_3\gamma) \cdot \frac{\phi[\{ln(y_{LB})-x_5\delta_{LB}\}/\sigma_{LB}]}{y_{LB}\sigma_{LB}} \\ & +\gamma_j \cdot \Phi(x_1\alpha) \cdot \Phi(x_2\beta) \cdot \phi(x_3\gamma) \cdot \frac{\phi[\{ln(y_{LB})-x_5\delta_{LB}\}/\sigma_{LB}]}{y_{LB}\sigma_{LB}} \\ & +\delta_{LB,j} \cdot \Phi(x_1\alpha) \cdot \Phi(x_2\beta) \cdot \Phi(x_3\gamma) \cdot \frac{\phi[\{ln(y_{LB})-x_5\delta_{LB}\}/\sigma_{LB}]}{y_{LB}\sigma_{LB}}\end{aligned} \tag{2}$$

To circumvent the possible problems associated with correlation between the error terms of the various equations integrated into equation (1), all of the average partial effects will be reported using standard errors bootstrapped from 200 replications of the entire multi-stage-model estimation procedure.

*4.2 Comments on control variables*

There are a few things to consider regarding the probit model specifications. First, one might reasonably argue that a farm's own size is endogenous, which would be problematic if the key variables for share of land under MSF cultivation were computed using our own sample. To bolster the causal interpretation of these parameters, the key variables are computed using CSO and MAL Crop Forecast Survey data, which are collected from more than 13,000 farmers annually and are representative at the district level.

Another modeling consideration is variables that may *not* belong in our model. That is, while the key variables, the share of land under control of larger farms, certainly belong in each stage of the model, it may seem sensible to control for other potentially confounding factors. In this regard, however, we exercise caution, because many of these "confounding factors" could be the mechanisms through which we expect the relationship of interest to be derived. For example, in

the stage of the model where the dependent variable is whether a smallholder sells to an LB, it might seem natural to think prices and other transaction-specific characteristics should be included, but this could be a mistake. Our model is an examination of whether the presence of larger farms increases the likelihood of even the smallest commercial farms selling to LBs. One reason we are testing this hypothesis is because of the transaction-specific differences between selling to LSTs versus smaller traders. Variables such as price do not belong in our probit model, because the *ceteris paribus* interpretation of the other coefficient estimates wouldn't make sense if they were included – why would a farmer choose to sell to an LST instead of a small trader if the prices (and everything else) were no different?[9] A truly *ceteris paribus* relationship between trader choice and our key variables wouldn't answer our research questions.[10]

That said, we are interested in the robustness of our results with respect to control variables. To that end, while we will use the relatively parsimonious model as the basis for analysis, we also present results from two versions with more control variables for comparison. The second model controls for household characteristics (age and gender of the household head, and the household size in adult equivalents) weather and climate variables (millimeters (mm) of main season rainfall, the 10-year average main season rainfall (mm), the ten-year average number of main season rain stress periods -defined as the overlapping number of 20 day periods in the growing season with less than 40 mm rainfall- and the 10-year coefficient of variation for rainfall)[11], and district-level transaction cost characteristics (median distance to tarmac roads (km), median distance traveled to an SST transactions (km), and the mean real cost of transporting maize (Kwacha/kg/km) to the point of SST sales). The third model, in addition to all aforementioned covariates, controls for provincial fixed effects, and allows for province-level intercept shifters for each time period. In the main body of the paper we will present and focus on just the partial effects of the farmland concentration variables, but a full set of coefficient estimates is available in the on-line appendix.

To summarize, we examine the relationship between prevailing farm size, LSTs, and small-scale farmers' maize market participation employing the following analytical steps. First we estimate a four stage model of market participation using probit specifications to test the effect of farm size consolidation on 1) the likelihood of selling maize, 2) the conditional likelihood of maize sellers selling to the private sector and 3) the conditional likelihood of those selling to the private sector selling to an LB, and in the fourth stage we estimate two lognormal models to examine the effects of land consolidation on the expected quantities sold to LBs and SBs. If LBs have an increasing presence in rural areas following the growth of MSF, we hypothesize that small-scale farmers might also benefit from the introduction of the relatively new marketing channel, which may offer better market access conditions as well as benefits from greater competition in the local market.

---

[9] Wooldridge (2003, p 200) discusses the mistake of including "too many" relevant variables using the example of estimating the effects of an alcohol tax on drunk driving fatalities. In that example, he points out, it would be an error to hold per capita alcohol consumption constant because reducing consumption is the mechanism by which a tax increase can be expected to reduce drunk driving.

[10] It may be worth further noting that controlling for some of these mechanisms could have the added problem of introducing regressors that are simultaneously determined.

[11] The source of weather & climate variables is TAMSAT (https://www.tamsat.org.uk/) and we are grateful to Nicole Mason and Jason Snyder for compiling these data.

## 5. Results

Table 3 shows the average partial effect estimates of farm-land consolidation on the various stages of the multi-stage model of maize market participation amongst farms cultivating less than 5 ha. The results for model (i) are the average partial effects (APE) at each stage of the model and bootstrapped standard errors as well as, in the last two columns, the APE on the "unconditional" expected values of sales to SBs and LBs. The model (i) results are computed using estimates of relatively simple models in each stage, where the only right-hand-side variables are the share of smallholder farmland within the observation's district that is cultivated on farms between 5 and 10 ha (B-farms), and the share of smallholder farmland cultivated on farms between 10 and 20 ha (C-farms). The results for model (ii) also show only the APEs for the farmland concentration variables, but in the underlying regressions for this model (full results in the on-line appendix) include control variables for household characteristics, weather and climate variables, and transaction cost characteristics as described in the conceptual framework. The results for model (iii) are again analogous to earlier results, but in this model the underlying regressions include all of the control variables from model (ii) plus controls for provincial effects within each time period (i.e., a dummy variable for each province as well as a dummy variable for each province interacted with a dummy variable for the 2015 time period with the necessary exclusion restrictions imposed to avoid perfect collinearity). Again, our discussion will focus on model (i), whereas models (ii) and (iii) allow us to examine the robustness of the results in model (i) with their full results available on-line.

In the first stage of the model (the first column, where n=15,087) we estimate a positive and statistically significant effect of farmland consolidation on the likelihood of a smaller farm selling maize. Interestingly, this effect is roughly double in magnitude given an increase in land under cultivation on farms between 10 and 20 ha. Both effects are also significant in regressions that include more control variables, though the magnitude of the C-farm effect is greater in model ii and that of B-farms is greater in model iii.[12]

The second stage (second column, n=7,215) shows that, conditional on the small farm being a maize seller, there is again positive correlation between farmland concentration and the likelihood of selling maize to the private sector, and again the magnitude of the effect is greater (and statistically significant) for C-farms. The estimated C-farm effect is still fairly large and significant after controlling for factors in model (ii). Interestingly, in the same model (ii) the estimated effect of B-farm concentration is greater than in model (i), and statistically significant at the 5% level. Neither effect is statistically significant in model (iii), but again, this is not

[12] It is difficult to articulate precisely what this means, because the difference (subtracting results of model (ii) from those of model (iii), say) is the product of the correlation between net provincial effects and the likelihood of selling and the net correlation between provincial effects and farmland concentration, all while holding weather, climate and transaction cost characteristics constant. For example, if the conditional correlation between net provincial effects and selling is positive, the fact that the estimated effect of B-farm concentration increases from model (ii) to model (iii) would tell us the net conditional correlation between provincial effects and B-farmland concentration is negative (and vice versa for C-farms, because that estimate decreases when provincial effects are included). However, this would not preclude the possibility of some provinces being positively (negatively) correlated with either B-farm (C-farm) concentration, other factors constant. Unfortunately, any more sensible or nuanced decomposition of what is driving this slightly anomalous result would require a much lengthier discussion and a much more specific model of the unobserved effects captured in the provincial variables in model (iii).

terribly surprising since this model holds constant all of the factors we expect to change when farm concentration increases, including the factors like distance to sellers and transaction costs that we believe act as the mechanisms of market evolution that are driving the correlations we observe in model (i).

In the third stage (third column, n=3,285) we estimate that amongst the farms less than 5 ha that sell to the private sector, those in districts where there is greater concentration of B-farms and C-farms are also more likely to sell to LBs, and the magnitude of the effects are similar, though the B-farm effect is more statistically significant. Similar to the first stage, though, the estimated C-farm effect is greater and more significant when controlling for factors in model (ii), whereas the B-farm effect is more pronounced in model (iii).

Results from the fourth stage of the model for the conditional expected sales to SBs (fourth column, n=2,784) shows that within the group of households that sell to *small-scale traders*, the expected sales are statistically significantly higher in districts where there is more farmland concentrated on B-farms (there is also a positive estimated effect from C-farms, but with a p-value of 0.14). At first glance, results from this stage may seem counter to the hypothesis that larger farms attract large-scale traders; a hypothesis seemingly supported by earlier studies and results described in our introduction. Recall, however, one of the important characteristics of the evolving markets we described is an increase in satellite buyers who act as aggregators for larger scale traders. In these data, such aggregators would be recorded as SSTs, which implies the results in column 4 could indeed be consistent with our hypothesis on the role of farmland consolidation in opening up markets to smaller farms. Notably these findings are largely robust across model specifications, with lower estimate variance in model (ii). The only exception is the C-farm effect in model (iii), which is not statistically significant.

Interestingly, although a rising presence of larger farms clearly suggests the smaller farms are more likely to sell and sell to LSTs and millers, the expected quantity of sales to LBs (given that a sale takes place) is not significantly greater than it otherwise would be (fifth column, n=501). In model (ii) there are statistically significant results, but they do not tell a consistent story, showing a positive effect from C-farms and a negative effect from B-farms. Model (iii) results, are again similar to model (i).

The average partial effects of farm consolidation on the unconditional expected sales (or expected per-household sales) to SBs and LBs are reported in the sixth and seventh columns respectively (n=15,087). To interpret these results we require a bit more context. For example, the table suggests that if the share of land under B-farms were to increase from 0 to 1, the expected share per household sales to SBs for farms less than 5 ha would increase by 343.6 kgs, but the example isn't actually possible (because there would be no small farms left). More realistically, if the share of land under B-farms were to increase from 0 (the actual minimum) to 0.101 (the median across districts and time), expected sales across all farms smaller than 5 ha to SBs would increase by roughly 35 kgs (a 20.4% increase compared to the current mean quantity of sales to SBs)[13]. Similarly, if the share of land under C-farms were to increase from 0 (the actual minimum) to 0.021 (the median across districts and time), expected sales across all farms smaller than 5 ha to SBs would increase by roughly 13 kgs (a 7.6% increase compared to the

[13] Mean sales across A-farms to SBs is 175.09 kg; to LBs is 85.38.

current mean quantity of sales to SBs). The expected per household increase in sales to LBs given the same change in C-farm concentration is another 9 kgs (a 10.5% increase compared to the current mean quantity sold to LBs). The effect of sales to LBs per household from an increase in B-farm concentration is also positive but not statistically significant. The trends in these results are similar in model (ii), and though all effects are positive in model (iii), only the B-farm concentration effect on smaller farm sales to SBs is statistically significant. Once again, this is not surprising because models (ii) and (iii) hold constant the factors we expect to change when farmland concentration increases, making sales to traders more attractive.

All together, the results from the multi-stage analysis imply smaller farmers in areas where more cultivated land can be found on farms between 5 and 20 hectares are more likely to sell maize, more likely to sell to the private sector, and more likely to sell to LBs. As a result, the expected sales to LBs across all farms smaller than 5 ha are slightly higher, but there is an even greater expected increase sales due to the increased presence of SSTs, which may be acting as aggregators.

## 6. Conclusions and Implications for Policy

This study is motivated by the need to better understand how the rise of medium-scale farms in Africa is affecting the welfare of small-scale farm households. Although survey evidence over the past decade has shown a recent and dramatic rise in the prevalence of "medium" sized farms between 5 and 100 hectares, smaller farms still constitute the vast majority of farms and rural households. Prior evidence highlights a co-evolution between the concentration of landholdings, surplus production volumes, and entry of larger traders into the market. The reasons for the rise in mid-sized farms and large-scale traders have been the subject of much discussion, but the aim of this study is to examine the impact of these changes.

It is not difficult to imagine how the rise of medium scale farms and large-scale traders could be detrimental to smaller farms: pushing less influential farmers to marginal lands, adding pressure to farmlands that reduce the frequency of fallow periods, or introducing quality and consistency standards that the smallest farmers may struggle to meet. It is also not obvious how a rise in larger-scale trading affects local prices of food, or whether changing food prices would be uniformly "good" or "bad" for surplus producing households. These are all questions worthy of and requiring more elaborate treatment than permitted by the scope of this study.

However, it is important to acknowledge that less obvious positive spill-over effects of farmland consolidation are also possible, and to investigate them. Large-scale traders, enticed by the presence of larger farms logically have lower unit-costs of transactions, and if they are competitive they may share these gains with any farmer selling to them. It is possible that the presence of larger farms opens entirely new marketing channels to their smaller neighbours. Indeed, what we call "medium sized farms" are somewhat akin to what Mellor (2017, p 75) calls "small commercial farms" that were harbingers of economic transformation and relative prosperity in many Asian countries. Are they playing a similar role in Africa? Our particular interest is to understand how the maize marketing options available to smallholders Zambian maize farmers are being influenced by the rise of medium-scale farms.

Our conclusions, supported by the results from estimating a four-stage model of market participation, are that the growth of medium scale farms results in even smaller farms being more likely to sell maize, small maize sellers being more likely to sell to the private sector, small farms selling maize to the private sector being more likely to sell to large-scale traders, and sell greater quantities to both large and small-scale traders that may be acting as aggregators for larger traders. Compared to a district where no farms are greater than 5 ha, the per-household maize sales amongst the smallest farms to a private trader in a district with the median share of farmland on medium sized farms is expected to be more than 55 kg (22%) higher. Where farmland is more concentrated, this figure could be more than twice as high.

Our findings contribute to the literature on "spillover effects" from large to small farms (Deininger and Xia, 2016; Lay, Nolte and Sipangule, 2018). Consistent with earlier evidence, our study finds that even the smallest farms experience spillover benefits from a growing number of medium-scale farms in their locality by encouraging new types of private investments in commodity value chains that improve market access.

A rising share of marketed output purchased by large-scale traders and a commensurate decline in the market share of small-scale traders (or a change in these SST's options for re-sale) may represent a logical evolution of agricultural markets following the emergence of MSFs. There are certainly farm structure and demand conditions where this kind of evolution may not be expected to occur, but commodity value chains have evolved in this way in most other parts of the world for the major field crops exhibiting economies of scale in procurement and distribution. While there are sensible reasons for believing this poses additional challenges to smaller farmers and small traders, the efficiency gains of a competitive large-scale trading sector cannot be ignored when considering whether the evolution is beneficial to smaller farms on balance.

**References**

Anseeuw, W., Jayne, T.S., Kachule, R. and Kotsopoulos, J. 'The quiet rise of medium-scale farms in Malawi', *Land*, Vol. 5(3), (2016) pp. 19-42.

Burke, W.J., Myers, R.J. and Jayne, T.S. 'A triple-hurdle model of production and market participation in Kenya's dairy market', *American Journal of Agricultural Economics*, Vol. 97(4), (2015) pp.1227-1246.

Daily Nation Newspaper. *FRA reduces maize purchase.* (2018) May 1, 2018. Available at: https://www.pressreader.com/zambia/daily-nation-newspaper/20180501/281590946183969 (last accessed October 1, 2018).

Deininger, K. and Xia, F. 'Quantifying spillover effects from large land-based investment: The case of Mozambique', *World Development* Vol. 87*,* (2016) pp. 227-241.

Fafchamps, M. 'Networks, communities and markets in sub-Saharan Africa: Implications for firm growth and investment', *Journal of African Economies* Vol. 10(suppl_2), (2001) pp. 109-142.

Government of Zambia (GRZ). *Central Statistics Office 2012 Rural Agricultural Livelihoods Survey (RALS12) Small & Medium Scale Holdings Instruction Manual for Listing, Sample Selection and Largest Maize Field Data Collection.* (2012) Lusaka, Zambia.

Jayne, T.S., Chamberlin, J., Traub, L., Sitko, N., Muyanga, M., Yeboah, F.K., Anseeuw, W., Chapoto, A., Wineman, A., Nkonde, C. and Kachule, R. 'Africa's changing farm size distribution patterns: The rise of medium-scale farms', *Agricultural Economics* Vol. 47(S1), (2016) pp. 197-214.

Lay, J., Nolte K. and Sipangule, K. *Large-scale farms and fsallholders: Evidence from Zambia,* Working Paper No 310 (German Institute of Global and Area Studies: Kiel, 2018).

Megill, D.J. *Recommendations on sample design for post-harvest surveys in Zambia based on the 2000 census*, Working Paper No. 11 (Food Security Research Project: Lusaka, 2004).

Mellor, J.W. *Agricultural Development and Economic Transformation: Promoting Growth with Poverty Reduction* (Cham, Switzerland: Springer Nature Publishing, 2017)

Messerli, P., Giger, M., Dwyer, M.B., Breu, T. and Eckert, S. 'The geography of large-scale land acquisitions: Analysing socio-ecological patterns of target contexts in the global South', *Applied Geography* Vol. 53(0), (2014) pp. 449-459.

Nolte, K. and Ostermeier, M. 'Labour market effects of large-scale agricultural investment: Conceptual considerations and estimated employment effects', *World Development* Vol. 98, (2017) pp. 430-446.

Nolte, K. and Sipangule, K. *Land use competition in sub-Saharan Africa's rural areas*, Policy Brief, No. 10/2017 (Poverty Reduction, Equity and Growth Network (PEGNet), Kiel Institute for the World Economy: Kiel, 2017).

Sitko, N. J., Burke, W.J. and Jayne, T.S. 'The quiet rise of large-scale trading firms in East and Southern Africa', *The Journal of Development Studies* Vol. 54(5), (2018) pp. 895-914.

Thurow, R. (2010). The Fertile Continent-Africa, Agriculture's Final Frontier. *Foreign Affairs*, *89*, 102.

Wooldridge, J. *Introductory Econometrics: A Modern Approach, 2e.* (Mason, Ohio: Thomson Learning, 2003).

**Figure 1. Multi-stage model of maize market participation for small farms**


Sources: Rural Agricultural Livelihood Surveys (2012, 2015) Note: LB is "Large buyer" and includes LSTs and millers, the former accounting for 67% of transactions and 60% of maize sold. Non-LB private transactions are primarily (over 75% of maize sold) with SSTs, but also includes sales to other households and otherwise uncategorized private transactions. There were 48 "A" farms that sold to both LB and SB groups. Strictly speaking, the hurdle framework only allows these to be included in the LB group in the 4th stage (counting only the quantity sold to LBs). Including these 48 observations in the non-LB quantity regression would not meaningfully affect results.

**Table 1: Maize marketing activity by farm size categories**

| Farm category (defined by area cultivated) | Share of farmers | Share of group that sell maize | Share of sellers that sell to private sector[a] | Share of sellers to private sector who sell to large-scale trader (LST) |
|---|---|---|---|---|
| "A"- farm <5 ha | 95% | 43% | 50% | 14% |
| "B"- 5 – 10 ha | 4% | 84% | 44% | 35% |
| "C"- 10 – 20 ha | 1% | 89% | 53% | 61% |

Source: Indaba Agricultural Policy Research Institute; Rural Agricultural Livelihoods Surveys, 2012 & 2015.

Note: a-The alternative to selling to the private sector is selling to the government.

**Table 2: Sample means over time for relevant variables in Zambia**

| | 2012 (n=8838) | 2015 (n=7933) |
|---|---|---|
| Sold maize (1=yes) | 0.52 | 0.51 |
| Sold to private sector (1=yes) | 0.19 | 0.27 |
| Sold to large-scale trader (1=yes | 0.02 | 0.06 |
| Kgs of maize sold | 1906.3 | 2035.4 |
| *Share of land cultivated on farms:*[a] | | |
| Less than 5 ha | 0.83 | 0.82 |
| 5 to 10 ha | 0.13 | 0.13 |
| 10 to 20 ha | 0.04 | 0.04 |
| Age of household head (years) | 45.49 | 48.66 |
| Education of household head (years) | 6.17 | 5.98 |
| Adult equivalents | 4.72 | 5.03 |
| Female household head (1=yes) | 0.19 | 0.21 |
| Main season rainfall (mm) | 795.0 | 838.5 |
| 10-year mean rainfall (mm) | 822.2 | 812.9 |
| 10-year mean number of rain stress periods | 0.74 | 0.80 |
| 10-year coefficient of rainfall variation | 11.47 | 9.96 |
| Distance to tarmac road (km) | 25.29 | 22.43 |
| Distance to small-scale trader transactions (km)[b] | 1.71 | 1.53 |
| Mean real transport cost to small-scale trader (Kwacha/kg/km)[b,c] | 0.03 | 0.04 |

Sources: IAPRI Rural Agricultural Livelihoods panel surveys (2012 & 2015)

Notes: a) District-level shares calculated from Central Statistics Office Crop Forecast Surveys (2012 & 2015); b) Means are an aggregate of transaction-level data-district medians amongst relevant observations are applied to all observations in the district; c) All prices are deflated to 2010 values in local currencies using IMF monthly consumer price indices available at data.imf.org.

**Table 3: Partial effect estimates of farmland consolidation from a multi-stage model of maize market participation for farms less than 5 hectares**

| | $\partial$Pr(sell)/ $\partial x$ | $\partial$Pr(*sell to private* \| *sell=1*)/ $\partial x$ | $\partial$Pr(*sell to LST or miller* \| *sell=1, sell to private=1*)/ $\partial x$ | $\partial$E(*Qsold* \| *sell=1, sell to private=1, sell to SB=1*)/ $\partial x$ | $\partial$E(*Qsold* \| *sell*=1*, sell to private*=1*, sell to LB=1*)/ $\partial x$ | $\partial$E(*sales to SST or other hh*)/ $\partial x$ | $\partial$E(*sales to LST or miller*)/ $\partial x$ |
|---|---|---|---|---|---|---|---|
| | n=15,087 | n=7,215 | n=3,285 | n=2,784 | n=501 | n=15,087 | n=15,087 |
| Model (i)-Farmland concentration only as explanatory variables | | | | | | | |
| District share of land | 0.23*** | 0.10 | 0.23** | 1,486.6*** | -3,160.9 | 343.6*** | 66.1 |
| under "B" farms | (0.08) | (0.13) | (0.10) | (341.0) | (2,289.5) | (83.7) | (55.6) |
| District share of land | 0.44*** | 1.24*** | 0.20 | 744.4* | 6,136.9* | 637.5*** | 411.0*** |
| under "C" farms | (0.11) | (0.19) | (0.14) | (447.7) | (3,383.5) | (108.0) | (87.3) |
| Model (ii)-Controlling for household characteristics: weather & climate; transaction cost characteristics | | | | | | | |
| District share of land | 0.12 | 0.29** | 0.06 | 1,109.3*** | -7,189.0** | 323.6*** | -84.5 |
| under "B" farms | (0.09) | (0.14) | (0.11) | (405.8) | (3,020.0) | (98.1) | (86.7) |
| District share of land | 0.67*** | 0.98*** | 0.57*** | 1,645.9*** | 7,567.6** | 740.2*** | 661.5*** |
| under "C" farms | (0.12) | (0.18) | (0.16) | (505.4) | (3,776.7) | (128.0) | (112.2) |
| Model (iii)-Controlling for hh characteristics: weather & climate; transaction costs characteristics; provincial, time & province*time effects | | | | | | | |
| District share of land | 0.64*** | 0.12 | 0.36** | 1,300.2** | -6,709.5* | 424.5*** | 112.6 |
| under "B" farms | (0.11) | (0.15) | (0.18) | (466.4) | (3,716.8) | (132.0) | (119.1) |
| District share of land | 0.15 | 0.02 | -0.07 | -253.1 | 4,212. | 28.8 | 95.3 |
| under "C" farms | (0.12) | (0.20) | (0.18) | (549.0) | (4,764.2) | (130.0) | (147.0) |

Sources: Household sales data from the Rural Agricultural Livelihood Surveys (2012, 2015); District farmland concentration variables from the Crop Forecast Surveys (2012; 2015). Notes: Bootstrapped standard errors from 200 replications in parentheses, *, **, *** indicates statistical significance at the 1, 5 and 10% levels respectively. Consistent with the definitions used by Zambia's Ministry of Agriculture and Livestock, "B-farms" are those cultivating between 5 and 10 ha and "C-farms" are those cultivating between 10 and 20 ha.

**Do medium-scale farms improve market access conditions for Zambian smallholders?**

## Appendix A: Some comments on the robustness to trader identification errors

Several early readers, reviewers, and participants at a session of the International Conference of Agricultural Economists rightly expressed concern regarding the reliance on farmers to correctly identify traders as either "large" or "small". Our primary response is to emphasize that we do believe the data are reasonably accurate for reasons discussed in the main text.

That said, we should certainly still consider the implications the possible errors a farmer could be making when they identify traders and examine how they may affect results. The first point is to point out that misidentifying traders is tantamount to dependent variable measurement error that is only relevant in the latter two stages of the four-hurdle model (i.e., misidentifying traders will not affect results in the probit analysis for 1) whether a farmer sells, or 2) whether they sell to the private sector). The implications of measurement error are exclusively relevant in stages 3 and 4. Moreover, whatever the implications are, they almost certainly have the largest effect on stage 3 results, where whether the farmer that is selling to the private sector indeed sells to a "large" buyer or not. In stage 4 the dependent variable is the quantity sold, so misidentifying traders would affect the number of observations available for stage 4 regression, but wouldn't introduce bias <u>unless, possibly, the measurement error in the third stage dependent variable is also correlated with the dependent variable in the fourth stage</u>. So, when we discuss the implications of potential measurement error bias, we will be focusing primarily on the third stage. Note, however, we are only setting aside the potential problems in the fourth stage temporarily; the fourth stage will be included in robustness checking regressions. Also, we will come back to discuss the implications of not meeting the underlined condition above, which we will argue is tertiary and probably unimportant once we have addressed the possible problems in the third stage.

Secondly, even if we only consider the third stage results, trader misidentification errors only present a problem if they are systemically correlated with the explanatory variables, and particularly the key variables. If trader misidentifications are not correlated with the concentration of farmland on larger farms, the errors would reduce the precision of our estimates, but results would still be consistent (and probably unbiased). We refer to Wooldridge (2003, pages 302-304) and Wooldridge (2010, pages 76-78) for a comprehensive treatment of this topic.

If, on the other hand, misidentification is correlated with explanatory variables, and particularly if measurement errors are correlated with our key explanatory variables of farmland concentration under "B" farms and "C" farms, then our estimated spillover effect estimates would indeed be biased. The direction of bias is determined by which traders are misidentified and how those errors are correlated (positively or negatively) with the key variables. So, we can categorize possible errors into four types: 1) LSTs being identified as SSTs in positive correlation with farmland concentration, 2) LSTs being identified as SSTs in negative correlation with farmland concentration, 3) SSTs being identified as LSTs in positive correlation with farmland concentration, and 4) SSTs being identified as LSTs in negative correlation with farmland concentration.

Table A1 summarizes the types of errors we could see, some examples of why we might see these errors, and their implications for the estimated spillover effects $(\hat{\beta})$ in a 2x2 matrix. We will henceforth refer to these as Quadrant 1 (Q1), Q2, Q3 or Q4 errors. To flesh out just part of these examples, consider the upper left quadrant of the matrix in Table A1. Taken to the extreme, if all farmers call all traders SSTs in areas where there are no large farms (even though some may have sold to what we would call an LST, LSTs would be artificially, systematically underrepresented in the data in areas where small farms prevail.

Correspondingly, the estimated "effect" of large farm presence on the likelihood of selling to an LST would be overstated (i.e., we would be more likely to find LSTs in the data in areas where farmland is concentrated, even if the actual distribution of LSTs was geographically uniform). As we expect the effect of farm consolidation to have a positive impact on the likelihood of selling to an LST, this would be bias *away* from zero. On the other hand, if farmers in areas of higher concentration understate the role of LSTs (e.g., because they consider some of what we consider LSTs "small" by comparison to others nearby traders, as in quadrant 2 of Table A1), the estimated effect of farm consolidation on the likelihood of selling to an LST would be underestimated, or biased *towards* zero. There are similar countervailing potential biases, depending on how measurement error is correlated with farmland concentration, if farmers were more likely to misidentify SSTs as LSTs (quadrants 3 & 4 of Table A1).

To us, there is no obvious reason to expect one kind of error more often than the other, nor one kind of correlation more often than the other. So, even though may seem naïve to believe there is no error in the farmer-defined trader characterization, we believe it is likely that such errors largely counterbalance each other. The additional noise from these errors almost certainly reduces the precision of our estimates, but may not cause any major systemic bias.

Nevertheless, we can examine the effects of each of the four types of problems in Table A1 as a robustness check by "correcting" each type of error, then examine how it affects results. That is, we can recode some of our observations between LSTs and SSTs in accordance with each different error type, then re-estimate the four-stage model with the "corrected" data. We can repeat this process many times in order to more fully understand how different these estimates are from those reported in our paper (i.e., we "bootstrap" the corrections to obtain standard errors for the "corrected" data results). We choose to do this 120 times, because that happens to be the highest number less than infinity in the table of critical z-score values we used to compute the 5$^{th}$ and 95$^{th}$ percentiles of "corrected" estimates (Wooldridge 2003, page 817).

Our method to "correct" the data is to recode a portion of observations from "small" to "large" buyers to examine potential influence of Q1 and Q2 type errors, and vice versa for Q3 and Q4 type errors. These recodes are done randomly, but with a probability proportional to, and either negatively (for Q1 and Q3 type errors) or positively (for Q2 and Q4 type errors) correlated with the concentration of farmland in the district. Specifically, for Q1 and Q3 errors, we set up the robustness check such that $\Pr(recode) = \max\,[0, \psi - (Bshare + Cshare)]$, where $\psi$ is chosen so that, on average, each of the 120 replications recodes 50 observations. Conversely, for Q2 and Q4 we use $\Pr(recode) = \psi(Bshare + Cshare)$, where again, $\psi$ is chosen so that, on average, each of the 120 replications recodes 50 observations.

In all cases, this amounts to a +/- 10% difference in "large" buyers on average. For the reasons we described above, we believe this is a larger degree of misidentification than should realistically be expected in our data. In other words, if our conclusions are robust to these changes, it is quite reasonable to assume they are robust to farmer misidentification errors.

Full details are in notes of tables described below, but for example, for Q1 $\psi = .1044$, so in districts with no "B" or "C" farms each observation has a 10.44% chance of being recoded, and that decreases as farmland concentration increases.

Tables A2.1-2.4 present the mean, standard deviation and the 5$^{th}$ and 95$^{th}$ percentiles for the marginal effects on conditional probabilities and expected values in stages 3 and 4, as well as the marginal effect on the unconditional expected values from the 120 replications of robustness checking for "correcting" Q1-Q4 type errors respectively. Table A2.0 presents results from the manuscript in the same format for comparison, substituting the 95% confidence interval (CI) for each parameter estimate from our main paper's results in the place of percentiles. Results from stages 1 and 2 of the model are not shown, since these are not affected by the robustness checking simulations, though these results are also computed within each repetition of the simulation and used to compute marginal effects on the "unconditional" expected values in the shaded region.

As expected, after correcting for Q1 and Q4 type errors, all of the estimates for stage 3 of the four-hurdle model are closer to zero. This is expected because Q1 and Q4 type measurement errors would have meant the results in R2.0 were positively biased, away from zero. Also as expected, after correcting for Q2 and Q3 type errors, all of the estimates for stage 3 are farther away from zero. Importantly, however, all of the point estimates in stage 3 after "correcting" Q1-Q4 type errors are within the 95% CI from the main paper. In fact, virtually all of the 5$^{th}$ and 95$^{th}$ percentiles for stage 3 in Tables A2.1-2.4 are within the corresponding 95% CIs in Table A2.0 – the only exception being the lower limit of marginal effect of B farmland concentration in Table A2.1.

Similarly, "correcting" for Q1-Q4 type errors has virtually no meaningful impact on the fourth stage estimates. The only considerable difference is for the impact B farmland concentration on the conditional expected sales to large buyers in Table A2.1– the manuscript's main results estimate a negative but non-significant effect, whereas the results in Table A2.1 are positive at least 95% of the time (please note, the we are careful not to refer to CIs in Tables A2.1-2.4 to avoid confusion – the hypotheses being tested in these tables are regarding the parameter **estimates**, not the underlying parameters themselves, as in Table A2.0. See table notes for more details). While this does stand out as a difference between robustness checking and our main results, it is important to note that the results in Table A2.1 are actually more in line with our main conclusions – that positive spillover effects exist – than the results in Table A2.0.

Perhaps most tellingly, the estimated effects of both B farmland and C farmland concentration on the unconditional expected value of sales to either large or small buyers are very similar in Tables A2.0-A2.4. All together, these tables show that our main findings appear to be highly robust to any reasonable expected measurement error that would be introducing bias to our main results.

At last, recall the earlier issue of whether trader misidentification should be much of a concern for the fourth stage result, which we've said is only a concern if "the measurement error in the third stage dependent variable is also correlated with the dependent variable in the fourth stage." In notational form, if $e$ is the measurement error in the third stage dependent variable and $Q_{sold}$ is the quantity sold, this condition states $Corr(Q_{sold}, e) = 0$. In fact, even if this condition were not met, the effects of violating it may have been captured in the robustness checking in Tables A2.0-2.4. That is because, after "correcting" the errors, what remains in the error term of stage 4 is not $e$, but $E(e|\psi, Bshare, Cshare)$. So, the only way stage 4 results could still be affected in the robustness checking results is if 1) some part of $e$ is correlated with $Q_{sold}$ after controlling for land shares, $Corr[Q_{sold}, E(e|\psi, Bshare, Cshare)] \neq 0$, and, 2) some part of $E(e|\psi, Bshare, Cshare)$ is correlated with farmland concentration, $Corr[(Bshare, Cshare), E(e|\psi, Bshare, Cshare)] \neq 0$ (because endogeneity is only a problem the error term is correlated with both the dependent and explanatory variables). If we are right about $\psi$, then the second inequality fails by definition. Also, recall we "estimated" $\psi$ to ensure the probability of error is 10% on average (more accurately, the appropriate value for $\psi$ is "found" using an iterative search program written in Stata). For reasons stated earlier, we think this is a larger degree of misidentification than should realistically be expected in our data, so any improvement on the estimated $\psi$ would almost certainly lead to results more similar to those in Table A2.0 than the results in A2.1-A2.4. This is why we say concerns with the fourth stage results are tertiary and probably unimportant once we have addressed the possible problems in the third stage.

In summary, there are four main points related to what we agree is a potential weakness in this study: the reliance on farmers to identify traders. First, we believe misidentification is probably not as much of a problem as it seems, both because of the reliability of the farmers and enumerators, and because of the survey support staff that was in place during enumeration that is described in the main text. Second, if misidentification has occurred, but these errors were uncorrelated with farmland concentration, our key estimates will be less precise, but consistent (and probably unbiased). Third, if misidentification occurred, but Q1, Q2, Q3 and Q4 type correlated errors counterbalance each other, estimates will be less precise and possibly inconsistent, but the degree of inconsistency approaches zero as counteracting effects approach equality. And finally, even if misidentification is a problem and one type of error is dominantly occurring, Tables A2.0-2.4 show our results are robust and our key findings would not be meaningfully different, even if we assume, on average and in probability proportional to farmland concentration, as many as 10% of the large traders are misidentified.

**Additional reference for this appendix:**

Wooldridge, J.M. *Econometric Analysis of Cross Section and Panel Data, 2e*. (Cambridge, MA: MIT Press, 2010).

**Table A1: Implications of measurement errors and misidentifying LSTs vs. SSTs**

| Type of measurement error/ Trader misidentification | Correlation between error and key explanatory variable (farmland concentration): Negative: They are less likely to make error where there are large farms | Correlation between error and key explanatory variable (farmland concentration): Positive: Farmers are more likely to make error where there are large farms |
|---|---|---|
| | Examples and Implications for $E(\hat{\beta})$ | |
| Farmer identifies a trader is an SST, when in fact they are an LST (a negative measurement error on trader size) | Quadrant 1 Example: Farmer doesn't know trader size but tells us an LST is an SST because they assume traders are usually small in areas where there are no large farms. Implication: $E(\hat{\beta}) > \beta$, "correcting" errors would likely result in lower estimated effects | Quadrant 2 Example: Farmer's perception of "large" trader is skewed in areas with more large farms, so some of what we would consider LSTs are called SSTs. Implication: $E(\hat{\beta}) < \beta$, "correcting" errors would likely result in higher estimated effects |
| Farmer identifies a trader is an LST, when in fact they are an SST (a positive measurement error on trader size) | Quadrant 3 Example: Farmer in areas without large farms, or any of what we would call LSTs, consider some traders "large", even though we would consider them SSTs. Implication: $E(\hat{\beta}) < \beta$, "correcting" errors would likely result in higher estimated effects | Quadrant 4 Example: Farmer doesn't know trader size but tells us an SST is an LST because they assume traders are usually large in areas where there are more large farms. Implication: $E(\hat{\beta}) > \beta$, "correcting" errors would likely result in lower estimated effects |

**Table A2.0: Main paper results for comparison to robustness simulations**

| Parameters | Estimate | Std. Err. | 95% | CI |
|---|---|---|---|---|
| *Stage 3* | | | | |
| $\frac{\partial P(sell\ to\ LB\ \|sell = 1, private = 1)}{\partial Bshare}$ | 0.23** | 0.10 | 0.04 | 0.42 |
| $\frac{\partial P(sell\ to\ LB\ \|sell = 1, private = 1)}{\partial Cshare}$ | 0.20 | 0.14 | -0.07 | 0.47 |
| *Stage 4* | | | | |
| $\frac{\partial E(Qsold\ \|sell\ to\ LB = 0, sell = 1, private = 1)}{\partial Bshare}$ | 1,486.6*** | 341.0 | 818.3 | 2,155.0 |
| $\frac{\partial E(Qsold\ \|sell\ to\ LB = 0, sell = 1, private = 1)}{\partial Cshare}$ | 744.4* | 447.7 | -133.1 | 1,621.9 |
| $\frac{\partial E(Qsold\ \|sell\ to\ LB = 1, sell = 1, private = 1)}{\partial Bshare}$ | -3,160.9 | 2,289.5 | -7,648.1 | 1,326.4 |
| $\frac{\partial E(Qsold\ \|sell\ to\ LB = 1, sell = 1, private = 1)}{\partial Cshare}$ | 6,136.9* | 3,383.5 | -494.7 | 12,768.44 |
| *Marginal effects on the unconditional expected value of sales* | | | | |
| $\frac{\partial E(Sales\ to\ SBs\ )}{\partial Bshare}$ | 343.6*** | 83.7 | 179.6 | 507.5 |
| $\frac{\partial E(Sales\ to\ SBs\ )}{\partial Cshare}$ | 637.5*** | 108.8 | 425.8 | 849.1 |
| $\frac{\partial E(Sales\ to\ LBs\ )}{\partial Bshare}$ | 66.1 | 55.6 | -42.9 | 175.1 |
| $\frac{\partial E(Sales\ to\ LBs\ )}{\partial Cshare}$ | 411.0*** | 87.3 | 240.0 | 582.0 |

Notes: Standard errors are from 200 bootstrapped replications of the 4-hurdle model described in Figure 1 and equation 1. Units for changes in expected values are kilograms of maize sold. Results from stages 1 and 2 from the model are not shown, since these are not affected by the robustness checking simulations, though these results are used to compute marginal effects on the "unconditional" expected values in the shaded region.
*, **, *** indicates statistical significance at the 1, 5 and 10% levels respectively for $H_0: \beta = 0\ vs. H_1: \beta \neq 0$.

**Table A2.1: "Correcting quadrant 1 errors" [a] simulated robustness comparison**

| *Parameters* | Estimate | Std. Dev. | Percentile 5 | Percentile 95 |
|---|---|---|---|---|
| *Stage 3* | | | | |
| $\frac{\partial P(sell\ to\ LB\ \|sell = 1, private = 1)}{\partial Bshare}$ | 0.05 | 0.04 | -0.02 | 0.12 |
| $\frac{\partial P(sell\ to\ LB\ \|sell = 1, private = 1)}{\partial Cshare}$ | 0.17*** | 0.01 | 0.15 | 0.19 |
| *Stage 4* | | | | |
| $\frac{\partial E(Qsold\ \|sell\ to\ LB = 0, sell = 1, private = 1)}{\partial Bshare}$ | 1,475.3*** | 28.1 | 1,419.8 | 1,530.9 |
| $\frac{\partial E(Qsold\ \|sell\ to\ LB = 0, sell = 1, private = 1)}{\partial Cshare}$ | 745.7*** | 11.6 | 722.8 | 768.7 |
| $\frac{\partial E(Qsold\ \|sell\ to\ LB = 1, sell = 1, private = 1)}{\partial Bshare}$ | 1,575.5* | 810.5 | -29.3 | 3,180.2 |
| $\frac{\partial E(Qsold\ \|sell\ to\ LB = 1, sell = 1, private = 1)}{\partial Cshare}$ | 6,526.2*** | 221.3 | 6,088.0 | 6,964.4 |
| *Marginal effects on the unconditional expected value of sales* | | | | |
| $\frac{\partial E(Sales\ to\ SBs\ )}{\partial Bshare}$ | 367.5*** | 6.7 | 354.4 | 380.7 |
| $\frac{\partial E(Sales\ to\ SBs\ )}{\partial Cshare}$ | 633.4*** | 3.0 | 627.5 | 639.3 |
| $\frac{\partial E(Sales\ to\ LBs\ )}{\partial Bshare}$ | 105.0*** | 15.0 | 75.3 | 134.7 |
| $\frac{\partial E(Sales\ to\ LBs\ )}{\partial Cshare}$ | 443.2*** | 11.6 | 420.2 | 466.3 |

Notes: Standard errors are from 120 bootstrapped replications of the 4-hurdle model described in Figure 1 and equation 1. Units for changes in expected values are kilograms of maize sold. Results from stages 1 and 2 from the model are not shown, since these are not affected by the robustness checking simulations, though these results are used to compute marginal effects on the "unconditional" expected values in the shaded region.
a-"Quadrant 1 errors" are when farmers identified a trader as "small" when we would have called them "large", and these errors are negatively correlated with the key explanatory variables, as described in Table R1. In the simulations we **"fix"** these errors by randomly recoding some transactions from **"small" to "large"** buyers with a probability that is **negatively correlated** with the district-level **concentration of land** under farms larger than 5 ha. Specifically, $\Pr(recode) = \max\,[0, \psi - (Bshare + Cshare)]$, where $\psi$ is chosen so that, on average each of the 120 replications recodes 50 observations (**a 10% difference in "large" buyers**). For this table, $\psi = 0.1044$, and the number of changes per replication ranges from 30 to70 (6% to 14% difference in the number of "large" buyers).*, **, *** indicates statistical significance at the 1, 5 and 10% levels respectively for $H_0: \hat{\beta} = 0\ vs. H_1: \hat{\beta} \neq 0$.

**Table A2.2: "Correcting quadrant 2 errors" [a] simulated robustness comparison**

| *Parameters* | Estimate | Std. Dev. | Percentile 5 | Percentile 95 |
|---|---|---|---|---|
| *Stage 3* | | | | |
| $\frac{\partial P(sell\ to\ LB\ \|sell = 1, private = 1)}{\partial Bshare}$ | 0.31*** | 0.05 | 0.22 | 0.41 |
| $\frac{\partial P(sell\ to\ LB\ \|sell = 1, private = 1)}{\partial Cshare}$ | 0.27*** | 0.07 | 0.13 | 0.41 |
| *Stage 4* | | | | |
| $\frac{\partial E(Qsold\ \|sell\ to\ LB = 0, sell = 1, private = 1)}{\partial Bshare}$ | 1,481.9*** | 63.6 | 1,355.9 | 1,607.9 |
| $\frac{\partial E(Qsold\ \|sell\ to\ LB = 0, sell = 1, private = 1)}{\partial Cshare}$ | 752.7*** | 89.2 | 576.0 | 929.3 |
| $\frac{\partial E(Qsold\ \|sell\ to\ LB = 1, sell = 1, private = 1)}{\partial Bshare}$ | -2,891.3** | 1,119.9 | -5,108.7 | -674.0 |
| $\frac{\partial E(Qsold\ \|sell\ to\ LB = 1, sell = 1, private = 1)}{\partial Cshare}$ | 4,005.9** | 1,688.3 | 663.0 | 7,348.8 |
| *Marginal effects on the unconditional expected value of sales* | | | | |
| $\frac{\partial E(Sales\ to\ SBs\ )}{\partial Bshare}$ | 319.9*** | 13.9 | 292.5 | 347.3 |
| $\frac{\partial E(Sales\ to\ SBs\ )}{\partial Cshare}$ | 613.0*** | 20.4 | 572.5 | 653.4 |
| $\frac{\partial E(Sales\ to\ LBs\ )}{\partial Bshare}$ | 94.2*** | 24.6 | 45.4 | 143.0 |
| $\frac{\partial E(Sales\ to\ LBs\ )}{\partial Cshare}$ | 426.1*** | 38.0 | 350.9 | 501.3 |

Notes: Standard errors are from 120 bootstrapped replications of the 4-hurdle model described in Figure 1 and equation 1. Units for changes in expected values are kilograms of maize sold. Results from stages 1 and 2 from the model are not shown, since these are not affected by the robustness checking simulations, though these results are used to compute marginal effects on the "unconditional" expected values in the shaded region.
a-"Quadrant 2 errors" are when farmers identified a trader as "small" when we would have called them "large", and these errors are negatively correlated with the key explanatory variables, as described in Table R1. In the simulations we **"fix"** these errors by randomly recoding some transactions from **"small" to "large"** buyers with a probability that is **positively correlated with** the district-level **concentration of land** under farms larger than 5 ha. Specifically, $\Pr(recode) = \psi(Bshare + Cshare)$, where $\psi$ is chosen so that, on average each of the 120 replications recodes 50 observations (a **10% difference in "large" buyers**). For this table, $\psi = 0.1011$, and the number of changes per replication ranges from 32 to 68 (6.4% to 13.6% difference in the number of "large" buyers). *, **, *** indicates statistical significance at the 1, 5 and 10% levels respectively for $H_0: \hat{\beta} = 0\ vs. H_1: \hat{\beta} \neq 0$.

**Table A2.3: "Correcting quadrant 3 errors" [a] simulated robustness comparison**

| *Parameters* | Estimate | Std. Dev. | Percentile 5 | Percentile 95 |
|---|---|---|---|---|
| *Stage 3* | | | | |
| $\frac{\partial P(sell\ to\ LB\ \|sell = 1, private = 1)}{\partial Bshare}$ | 0.30*** | 0.03 | 0.25 | 0.35 |
| $\frac{\partial P(sell\ to\ LB\ \|sell = 1, private = 1)}{\partial Cshare}$ | 0.22*** | 0.02 | 0.18 | 0.26 |
| *Stage 4* | | | | |
| $\frac{\partial E(Qsold\ \|sell\ to\ LB = 0, sell = 1, private = 1)}{\partial Bshare}$ | 1,418.9*** | 29.4 | 1,360.6 | 1,477.2 |
| $\frac{\partial E(Qsold\ \|sell\ to\ LB = 0, sell = 1, private = 1)}{\partial Cshare}$ | 750.1*** | 32.9 | 685.1 | 815.2 |
| $\frac{\partial E(Qsold\ \|sell\ to\ LB = 1, sell = 1, private = 1)}{\partial Bshare}$ | -3,037.8*** | 633.5 | -4,292.2 | -1,783.5 |
| $\frac{\partial E(Qsold\ \|sell\ to\ LB = 1, sell = 1, private = 1)}{\partial Cshare}$ | 5,951.0*** | 385.0 | 5,188.8 | 6,713.3 |
| *Marginal effects on the unconditional expected value of sales* | | | | |
| $\frac{\partial E(Sales\ to\ SBs\ )}{\partial Bshare}$ | 324.8*** | 8.9 | 307.2 | 342.4 |
| $\frac{\partial E(Sales\ to\ SBs\ )}{\partial Cshare}$ | 654.9*** | 10.8 | 633.5 | 676.2 |
| $\frac{\partial E(Sales\ to\ LBs\ )}{\partial Bshare}$ | 100.7*** | 13.0 | 75.0 | 126.5 |
| $\frac{\partial E(Sales\ to\ LBs\ )}{\partial Cshare}$ | 383.1*** | 14.4 | 354.7 | 411.6 |

Notes: Standard errors are from 120 bootstrapped replications of the 4-hurdle model described in Figure 1 and equation 1. Units for changes in expected values are kilograms of maize sold. Results from stages 1 and 2 from the model are not shown, since these are not affected by the robustness checking simulations, though these results are used to compute marginal effects on the "unconditional" expected values in the shaded region.
a-"Quadrant 1 errors" are when farmers identified a trader as "small" when we would have called them "large", and these errors are negatively correlated with the key explanatory variables, as described in Table R1. In the simulations we **"fix"** these errors by randomly recoding some transactions from **"large" to "small"** buyers with a probability that is **negatively correlated** with the district-level **concentration of land** under farms larger than 5 ha. Specifically, $\Pr(recode) = \max\,[0, \psi - (Bshare + Cshare)]$, where $\psi$ is chosen so that, on average each of the 120 replications recodes 50 observations (**a 10% difference in "large" buyers**). For this table, $\psi = 0.2869$, and the number of changes per replication ranges from 35 to 69 (7% to 13.8% difference in the number of "large" buyers). *, **, *** indicates statistical significance at the 1, 5 and 10% levels respectively for $H_0: \hat{\beta} = 0\ vs. H_1: \hat{\beta} \neq 0$.

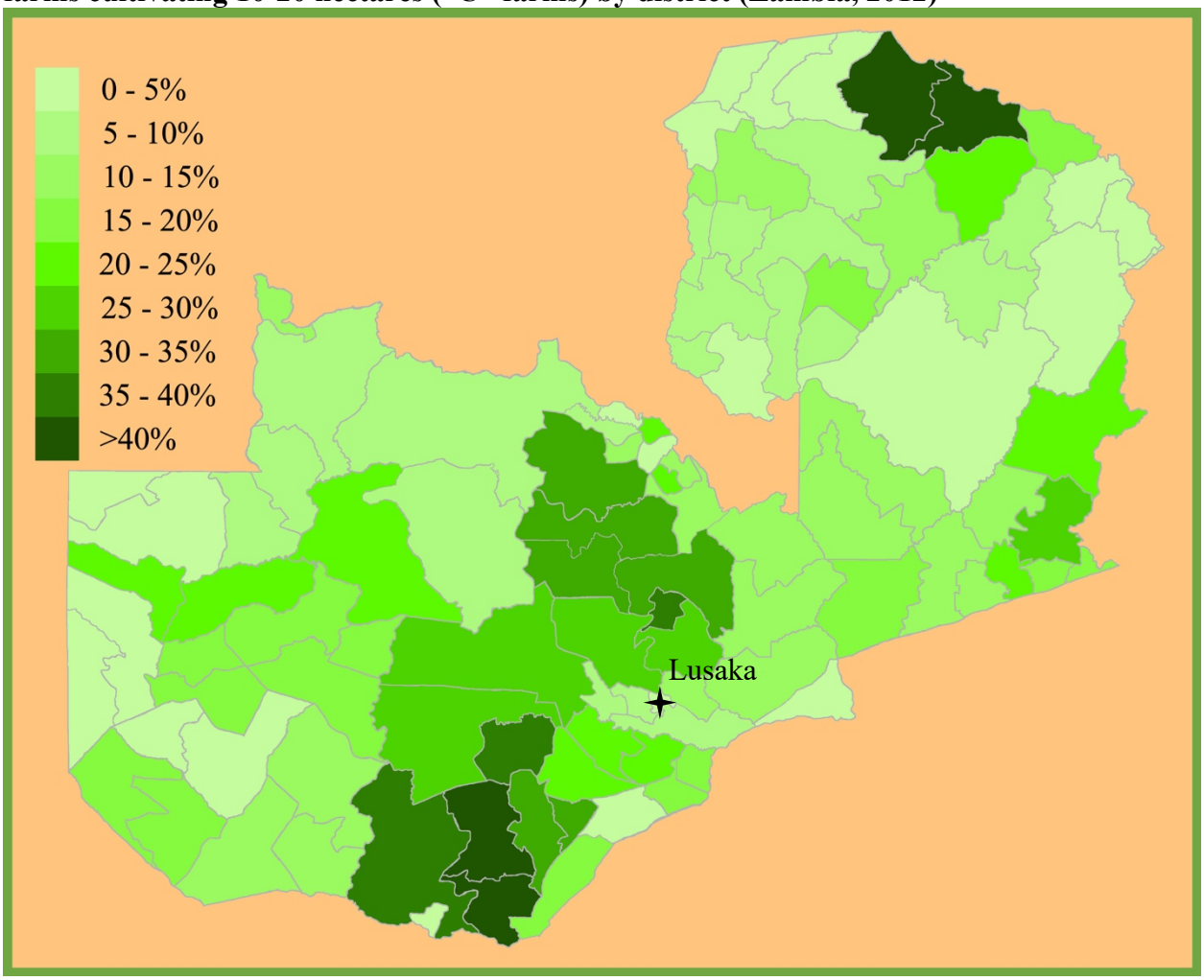
**Table A2.4: "Correcting quadrant 4 errors" [a] simulated robustness comparison**

| *Parameters* | Estimate | Std. Dev. | Percentile 5 | Percentile 95 |
|---|---|---|---|---|
| *Stage 3* | | | | |
| $\frac{\partial P(sell\ to\ LB\ \|sell = 1, private = 1)}{\partial Bshare}$ | 0.15*** | 0.04 | 0.07 | 0.23 |
| $\frac{\partial P(sell\ to\ LB\ \|sell = 1, private = 1)}{\partial Cshare}$ | 0.11** | 0.06 | 0.00 | 0.22 |
| *Stage 4* | | | | |
| $\frac{\partial E(Qsold\ \|sell\ to\ LB = 0, sell = 1, private = 1)}{\partial Bshare}$ | 1,554.7*** | 47.4 | 1,460.8 | 1,648.6 |
| $\frac{\partial E(Qsold\ \|sell\ to\ LB = 0, sell = 1, private = 1)}{\partial Cshare}$ | 884.1*** | 76.5 | 732.6 | 1,035.6 |
| $\frac{\partial E(Qsold\ \|sell\ to\ LB = 1, sell = 1, private = 1)}{\partial Bshare}$ | -3,209.9*** | 842.0 | -4,877.0 | -1,542.7 |
| $\frac{\partial E(Qsold\ \|sell\ to\ LB = 1, sell = 1, private = 1)}{\partial Cshare}$ | 6,200.7*** | 1,611.4 | 3,010.2 | 9,391.2 |
| *Marginal effects on the unconditional expected value of sales* | | | | |
| $\frac{\partial E(Sales\ to\ SBs\ )}{\partial Bshare}$ | 377.5*** | 14.5 | 348.8 | 406.3 |
| $\frac{\partial E(Sales\ to\ SBs\ )}{\partial Cshare}$ | 699.3*** | 22.9 | 653.9 | 744.7 |
| $\frac{\partial E(Sales\ to\ LBs\ )}{\partial Bshare}$ | 37.5* | 19.5 | -1.1 | 76.1 |
| $\frac{\partial E(Sales\ to\ LBs\ )}{\partial Cshare}$ | 345.0*** | 34.2 | 277.3 | 412.7 |

Notes: Standard errors are from 120 bootstrapped replications of the 4-hurdle model described in Figure 1 and equation 1. Units for changes in expected values are kilograms of maize sold. Results from stages 1 and 2 from the model are not shown, since these are not affected by the robustness checking simulations, though these results are used to compute marginal effects on the "unconditional" expected values in the shaded region.
a-"Quadrant 1 errors" are when farmers identified a trader as "small" when we would have called them "large", and these errors are negatively correlated with the key explanatory variables, as described in Table R1. In the simulations we **"fix"** these errors by randomly recoding some transactions from **"large" to "small"** buyers with a probability that is **positively correlated with** the district-level **concentration of land** under farms larger than 5 ha. Specifically, $\Pr(recode) = \psi(Bshare + Cshare)$, where $\psi$ is chosen so that, on average each of the 120 replications recodes 50 observations (a **10% difference in "large" buyers**). For this table, $\psi = 0.4733$, and the number of changes per replication ranges from 31 to 63 (6.2% to 12.6% difference in the number of "large" buyers). *, **, *** indicates statistical significance at the 1, 5 and 10% levels respectively for $H_0: \hat{\beta} = 0\ vs. H_1: \hat{\beta} \neq 0$.

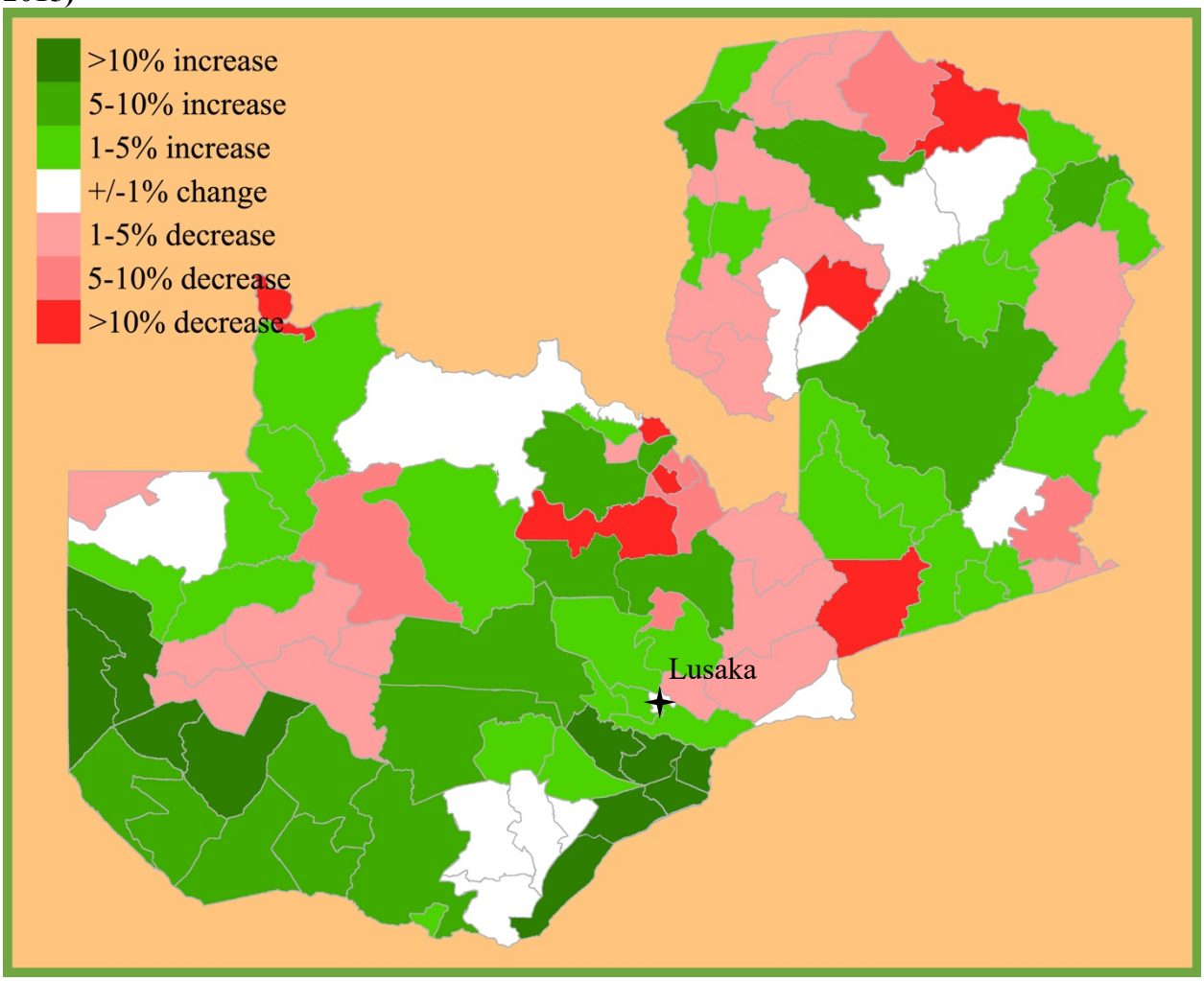
## Appendix B: Supporting figures

**Figure B1: Combined share of land under farms cultivating 5-10 hectares ("B" farms) and farms cultivating 10-20 hectares ("C" farms) by district (Zambia, 2012)**




Source: Crop Forecast Surveys (Zambia Central Statistics Office, 2012). Map of administrative borders adapted from OnTheWorldMap.com. Calculations are based on 2012 administrative boundaries. Consistent with the definitions used by Zambia's Ministry of Agriculture and Livestock, "B-farms" are those cultivating between 5 and 10 ha and "C-farms" are those cultivating between 10 and 20 ha.

**Figure B2. Percentage point change in combined share of land under "B" and "C" farms (2012-2015)**




Source: Crop Forecast Surveys (Zambia Central Statistics Office, 2012, 2015). Map of administrative borders adapted from OnTheWorldMap.com. Calculations are based on 2012 administrative boundaries. Consistent with the definitions used by Zambia's Ministry of Agriculture and Livestock, "B-farms" are those cultivating between 5 and 10 ha and "C-farms" are those cultivating between 10 and 20 ha.

## APPENDIX C: Select Stata (v.15.1) code and full results for 4-hurdle models

***Program used to compute 4-hurdle model bootstrapped standard errors:**

```
program define APEboot, rclass
 1. preserve
 2. probit sell $indep [pw=wei] if Afarm==1
 3. global a_B=_b[Bshare]
 4. global a_C=_b[Cshare]
 5. predict xa, xb
 6. margins, dydx($indep)
 7. return scalar dPsell_dB= el(r(b),1,1)
 8. return scalar dPsell_dC= el(r(b),1,2)
 9. probit private $indep [pw=wei] if Afarm==1 & sell==1
10. global b_B=_b[Bshare]
11. global b_C=_b[Cshare]
12. predict xb, xb
13. margins, dydx($indep)
14. return scalar dPpriv_dB= el(r(b),1,1)
15. return scalar dPpriv_dC= el(r(b),1,2)
16. probit lgmill $indep [pw=wei] if Afarm==1 & sell==1 & private==1
17. global g_B=_b[Bshare]
18. global g_C=_b[Cshare]
19. predict xg, xb
20. margins, dydx($indep)
21. return scalar dPlgmil_dB= el(r(b),1,1)
22. return scalar dPlgmil_dC= el(r(b),1,2)
23. lnormal SSTmzkg $indep [pw=wei] if Afarm==1 & sell==1 & private==1 & lgmill==0
24. global d1_B=_b[Bshare]
25. global d1_C=_b[Cshare]
26. global sig1=e(sigma)
27. predict xd1, xb
28. gen d1b=_b[Bshare]*exp(xd1+$sig1 * $sig1 /2) if e(sample)
29. gen d1c=_b[Cshare]*exp(xd1+$sig1 * $sig1 /2) if e(sample)
30. sum d1b
31. return scalar d1_B=r(mean)
32. sum d1c
33. return scalar d1_C=r(mean)
34. lnormal LSTmzkg $indep [pw=wei] if Afarm==1 & sell==1 & private==1 & lgmill==1
35. global d2_B=_b[Bshare]
36. global d2_C=_b[Cshare]
37. global sig2=e(sigma)
38. predict xd2, xb
39. gen d2b=_b[Bshare]*exp(xd2+$sig1 * $sig1 /2) if e(sample)
40. gen d2c=_b[Cshare]*exp(xd2+$sig1 * $sig1 /2) if e(sample)
41. sum d2b
42. return scalar d2_B=r(mean)
43. sum d2c
44. return scalar d2_C=r(mean)
45. ***********SST
      gen dQsst_dB=$a_B * normalden(xa)*normal(xb)*normal(-xg)*exp(xd1+$sig1 * $sig1 /2)  ///
             +$b_B * normal(xa)*normalden(xb)*normal(-xg)*exp(xd1+$sig1 * $sig1 /2)  ///
             -$g_B * normal(xa)*normal(xb)*normalden(-xg)*exp(xd1+$sig1 * $sig1 /2)  ///
             +$d1_B * normal(xa)*normal(xb)*normal(-xg)*exp(xd1+$sig1 * $sig1 /2)
46. sum dQsst_dB if Afarm==1
47. return scalar dQsst_dB=r(mean)
48. gen dQsst_dC=$a_C * normalden(xa)*normal(xb)*normal(-xg)*exp(xd1+$sig1 * $sig1 /2)  ///
            +$b_C * normal(xa)*normalden(xb)*normal(-xg)*exp(xd1+$sig1 * $sig1 /2)  ///
            -$g_C * normal(xa)*normal(xb)*normalden(-xg)*exp(xd1+$sig1 * $sig1 /2)  ///
            +$d1_C * normal(xa)*normal(xb)*normal(-xg)*exp(xd1+$sig1 * $sig1 /2)
49. sum dQsst_dC if Afarm==1
50. return scalar dQsst_dC=r(mean)
51. **********LST
      gen dQlst_dB=$a_B * normalden(xa)*normal(xb)*normal(xg)*exp(xd2+$sig2 * $sig2 /2)  ///
            +$b_B * normal(xa)*normalden(xb)*normal(xg)*exp(xd2+$sig2 * $sig2 /2)  ///
            +$g_B * normal(xa)*normal(xb)*normalden(xg)*exp(xd2+$sig2 * $sig2 /2)  ///
            +$d2_B * normal(xa)*normal(xb)*normal(xg)*exp(xd2+$sig2 * $sig2 /2)
52. sum dQlst_dB if Afarm==1
53. return scalar dQlst_dB=r(mean)
54. gen dQlst_dC=$a_C * normalden(xa)*normal(xb)*normal(xg)*exp(xd2+$sig2 * $sig2 /2)  ///
            +$b_C * normal(xa)*normalden(xb)*normal(xg)*exp(xd2+$sig2 * $sig2 /2)  ///
            +$g_C * normal(xa)*normal(xb)*normalden(xg)*exp(xd2+$sig2 * $sig2 /2)  ///
            +$d2_C * normal(xa)*normal(xb)*normal(xg)*exp(xd2+$sig2 * $sig2 /2)
55. sum dQlst_dC if Afarm==1
56. return scalar dQlst_dC=r(mean)
57. restore
58. end
```

***Description of variables**

| variable name | variable label |
|---|---|
| sell | Did hh sell maize (1=yes; 0=no) |
| private | Did hh sell maize to private sector buyer (1=yes; 0=no) |
| lgmill | Did hh sell maize to LST/miller (1=yes; 0=no) |
| SSTmzkg | Quantity sold to SST/SBs (kg) |
| LSTmzkg | Quantity sold to LST/miller (kg) |
| Bshare | Share of cultivated area under 'B' farms 5-10 ha (from CFS) |
| Cshare | Share of cultivated area under 'C' farms >10 ha (from CFS) |
| headage | Age of household head |
| eduhead | Level of education hh head in years |
| ae | Adult equivalents |
| fhead | Female headed household |
| rain | Growing season rainfall total |
| rain10 | Long-run avg growing season rainfall (10yr) |
| stress10 | Long-run avg stress periods (10yr) |
| cv10 | Long-run coef of rain var (10yr) |
| Dtarmac | District median distance to nearest tarmac (km) |
| DKMsstall | District median distance to SST |
| avRTCsst | District mean real (2010) transport cost to small trader (K/kg) |
| P1 | Province = 1 (1=yes; 0=no) |
| P2 | Province = 2 (1=yes; 0=no) |
| P3 | Province = 3 (1=yes; 0=no) |
| P4 | Province = 4 (1=yes; 0=no) |
| P5 | Province = 5 (1=yes; 0=no) |
| P6 | Province = 6 (1=yes; 0=no) |
| P7 | Province = 7 (1=yes; 0=no) |
| P8 | Province = 8 (1=yes; 0=no) |
| P9 | Province = 9 (1=yes; 0=no) |
| P10 | Province = 10 (1=yes; 0=no) |
| P1_15 | Province = 1 & year=2015 (1=yes; 0=no) |
| P2_15 | Province = 2 & year=2015 (1=yes; 0=no) |
| P3_15 | Province = 3 & year=2015 (1=yes; 0=no) |
| P4_15 | Province = 4 & year=2015 (1=yes; 0=no) |
| P5_15 | Province = 5 & year=2015 (1=yes; 0=no) |
| P6_15 | Province = 6 & year=2015 (1=yes; 0=no) |
| P7_15 | Province = 7 & year=2015 (1=yes; 0=no) |
| P8_15 | Province = 8 & year=2015 (1=yes; 0=no) |
| P9_15 | Province = 9 & year=2015 (1=yes; 0=no) |
| P10_15 | Province = 10 & year=2015 (1=yes; 0=no) |

```
*Model i) Main 4-hurdle model
global indep Bshare Cshare

* Model i: Stage 1
probit sell $indep [pw=wei] if Afarm==1
(Iteration log output omitted)

Probit regression                               Number of obs     =     15,087
                                                Wald chi2(2)      =      45.51
                                                Prob > chi2       =     0.0000
Log pseudolikelihood = -1893575.2               Pseudo R2         =     0.0041

------------------------------------------------------------------------------
             |               Robust
        sell |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      Bshare |   .5861528   .1935311     3.03   0.002     .2068387    .9654668
      Cshare |   1.134947   .2903311     3.91   0.000     .5659087    1.703986
       _cons |  -.3170321   .0252714   -12.55   0.000     -.366563   -.2675011
------------------------------------------------------------------------------


* Model i: Stage 2
probit private $indep [pw=wei] if Afarm==1 & sell==1
(Iteration log output omitted)

Probit regression                               Number of obs     =      7,215
                                                Wald chi2(2)      =      87.76
                                                Prob > chi2       =     0.0000
Log pseudolikelihood =  -808623.8               Pseudo R2         =     0.0166

------------------------------------------------------------------------------
             |               Robust
     private |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      Bshare |   .2675795    .288868     0.93   0.354    -.2985913    .8337503
      Cshare |   3.168791   .4188809     7.56   0.000     2.347799    3.989783
       _cons |  -.1887491   .0385729    -4.89   0.000    -.2643507   -.1131475
------------------------------------------------------------------------------


* Model i: Stage 3
probit lgmill $indep [pw=wei] if Afarm==1 & sell==1 & private==1
(Iteration log output omitted)

Probit regression                               Number of obs     =      3,285
                                                Wald chi2(2)      =      14.61
                                                Prob > chi2       =     0.0007
Log pseudolikelihood = -235276.47               Pseudo R2         =     0.0086

------------------------------------------------------------------------------
             |               Robust
      lgmill |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      Bshare |    1.04803   .4845699     2.16   0.031     .0982908     1.99777
      Cshare |     .90001   .6234097     1.44   0.149    -.3218506    2.121871
       _cons |    -1.3089   .0710065   -18.43   0.000    -1.448071    -1.16973
------------------------------------------------------------------------------
```

*** Model i: Stage 4: SB quantity regression**

```
lnormal SSTmzkg $indep [pw=wei] if Afarm==1 & sell==1 & private==1 & lgmill==0
(Iteration log output omitted)

                                                  Number of obs     =      2,784
                                                  Wald chi2(2)      =      41.60
Log pseudolikelihood = -3902861.6                 Prob > chi2       =     0.0000

------------------------------------------------------------------------------
             |               Robust
     SSTmzkg |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      Bshare |   1.768057   .3761726     4.70   0.000     1.030772    2.505342
      Cshare |   .8853269   .5647821     1.57   0.117    -.2216256    1.992279
       _cons |   5.769161   .0510376   113.04   0.000     5.669129    5.869192
-------------+----------------------------------------------------------------
    /lnsigma |   .1512168   .0181165     8.35   0.000      .115709    .1867245
-------------+----------------------------------------------------------------
       sigma |   1.163249    .021074                      1.122669    1.205295
------------------------------------------------------------------------------
```

*** Model i: Stage 4: LB quantity regression**

```
lnormal LSTmzkg $indep [pw=wei] if Afarm==1 & sell==1 & private==1 & lgmill==1
(Iteration log output omitted)

                                                  Number of obs     =        501
                                                  Wald chi2(2)      =       3.86
Log pseudolikelihood = -689876.78                 Prob > chi2       =     0.1449

------------------------------------------------------------------------------
             |               Robust
     LSTmzkg |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      Bshare |  -1.236611   .8755509    -1.41   0.158    -2.952659    .4794369
      Cshare |   2.400916   1.288302     1.86   0.062    -.1241094    4.925941
       _cons |    7.20374   .1376102    52.35   0.000     6.934029    7.473451
-------------+----------------------------------------------------------------
    /lnsigma |   -.063107   .0379396    -1.66   0.096    -.1374673    .0112532
-------------+----------------------------------------------------------------
       sigma |    .938843   .0356193                      .8715629    1.011317
------------------------------------------------------------------------------
```

```
* Model i: Bootstrap the full model to predict all partial effects
bootstrap         dPsell_dB=r(dPsell_dB)          ///
                  dPsell_dC=r(dPsell_dC)          ///
                  dPpriv_dB=r(dPpriv_dB)          ///
                  dPpriv_dC=r(dPpriv_dC)          ///
                  dPlgmil_dB=r(dPlgmil_dB)        ///
                  dPlgmil_dC=r(dPlgmil_dC)        ///
                  d1_B=r(d1_B)                            ///
                  d1_C=r(d1_C)                            ///
                  d2_B=r(d2_B)                            ///
                  d2_C=r(d2_C)                            ///
                  dQsst_dB=r(dQsst_dB)            ///
                  dQsst_dC=r(dQsst_dC)            ///
                  dQlst_dB=r(dQlst_dB)            ///
                  dQlst_dC=r(dQlst_dC)            ///
                  if Afarm==1, seed(39) reps(200): APEboot

(Iteration log output omitted)

Bootstrap results                               Number of obs     =     15,087
                                                Replications      =        200

------------------------------------------------------------------------------
             |   Observed   Bootstrap                         Normal-based
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   dPsell_dB |   .2287187   .0802253     2.85   0.004       .07148    .3859574
   dPsell_dC |   .4428601   .1113642     3.98   0.000     .2245903      .66113
   dPpriv_dB |   .1048435    .133198     0.79   0.431    -.1562199    .3659068
   dPpriv_dC |   1.241601   .1860644     6.67   0.000     .8769218    1.606281
  dPlgmil_dB |   .2277102   .0956825     2.38   0.017     .0401759    .4152445
  dPlgmil_dC |   .1955491   .1375619     1.42   0.155    -.0740673    .4651656
        d1_B |   1486.642    341.015     4.36   0.000     818.2644    2155.019
        d1_C |   744.4126   447.7077     1.66   0.096    -133.0783    1621.903
        d2_B |  -3160.854   2289.467    -1.38   0.167    -7648.126    1326.419
        d2_C |   6136.887   3383.506     1.81   0.070    -494.6631    12768.44
    dQsst_dB |   343.5773   83.65822     4.11   0.000     179.6102    507.5444
    dQsst_dC |   637.4755    107.997     5.90   0.000     425.8052    849.1459
    dQlst_dB |   66.07714   55.61436     1.19   0.235    -42.92501    175.0793
    dQlst_dC |   410.9939   87.26589     4.71   0.000     239.9559    582.0319
------------------------------------------------------------------------------
```

**Model ii) 4-hurdle model w/hh; weather & transaction vars**

```
global indep Bshare Cshare headage eduhead ae fhead rain rain10 stress10 cv10 Dtarmac DKMssta
avRTCsst
```

*** Model ii: Stage 1**

```
probit sell $indep [pw=wei] if Afarm==1
```

*(Iteration log output omitted)*

```
Probit regression                               Number of obs     =     15,082
                                                Wald chi2(13)     =     493.03
                                                Prob > chi2       =     0.0000
Log pseudolikelihood = -1825299.2               Pseudo R2         =     0.0397
------------------------------------------------------------------------------
             |               Robust
        sell |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      Bshare |   .3166716    .224613     1.41   0.159    -.1235618    .7569049
      Cshare |   1.797174    .313912     5.73   0.000     1.181918    2.412431
     headage |  -.0003138   .0009291    -0.34   0.736    -.0021349    .0015072
     eduhead |   .0421057   .0040506    10.39   0.000     .0341666    .0500448
          ae |   .0441288    .006771     6.52   0.000      .030858    .0573997
       fhead |  -.1274241   .0347186    -3.67   0.000    -.1954713    -.059377
        rain |  -.0009281    .000247    -3.76   0.000    -.0014122   -.0004439
      rain10 |  -.0011401   .0003404    -3.35   0.001    -.0018074   -.0004729
    stress10 |  -.1884456   .0351971    -5.35   0.000    -.2574306   -.1194606
        cv10 |  -.0218015   .0055505    -3.93   0.000    -.0326803   -.0109227
     Dtarmac |  -.0028398   .0005006    -5.67   0.000     -.003821   -.0018586
   DKMsstall |  -.0161748   .0024414    -6.63   0.000      -.02096   -.0113897
    avRTCsst |  -1.221185   .2803739    -4.36   0.000    -1.770707    -.671662
       _cons |   1.501759   .2998738     5.01   0.000     .9140171    2.089501
------------------------------------------------------------------------------
```

*** Model ii: Stage 2**

```
probit private $indep [pw=wei] if Afarm==1 & sell==1
```

*(Iteration log output omitted)*

```
Probit regression                               Number of obs     =      7,212
                                                Wald chi2(13)     =     175.47
                                                Prob > chi2       =     0.0000
Log pseudolikelihood = -793217.47               Pseudo R2         =     0.0349
------------------------------------------------------------------------------
             |               Robust
     private |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      Bshare |   .7496496   .3308667     2.27   0.023     .1011628    1.398136
      Cshare |   2.559655   .4357865     5.87   0.000     1.705529    3.413781
     headage |  -.0019444   .0014228    -1.37   0.172    -.0047331    .0008444
     eduhead |  -.0283962   .0058991    -4.81   0.000    -.0399583   -.0168341
          ae |  -.0381995   .0096503    -3.96   0.000    -.0571137   -.0192853
       fhead |   .0312309   .0543316     0.57   0.565     -.075257    .1377188
        rain |   .0021417   .0003778     5.67   0.000     .0014013    .0028821
      rain10 |  -.0004602   .0004969    -0.93   0.354     -.001434    .0005137
    stress10 |   .0216332   .0527738     0.41   0.682    -.0818016    .1250681
        cv10 |   .0231939   .0084659     2.74   0.006      .006601    .0397868
     Dtarmac |   .0009856   .0007991     1.23   0.217    -.0005806    .0025517
   DKMsstall |   .0087571   .0051356     1.71   0.088    -.0013085    .0188226
    avRTCsst |   1.546676   .7181351     2.15   0.031     .1391571    2.954195
       _cons |  -1.502373     .44684    -3.36   0.001    -2.378163   -.6265822
------------------------------------------------------------------------------
```

*** Model ii: Stage 3**

```
probit lgmill $indep [pw=wei] if Afarm==1 & sell==1 & private==1

(Iteration log output omitted)

Probit regression                                 Number of obs     =      3,285
                                                  Wald chi2(13)     =     107.61
                                                  Prob > chi2       =     0.0000
Log pseudolikelihood = -222189.83                 Pseudo R2         =     0.0637
------------------------------------------------------------------------------
             |               Robust
      lgmill |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      Bshare |   .3064541   .5933057     0.52   0.605    -.8564037    1.469312
      Cshare |   2.762789   .7104664     3.89   0.000       1.3703    4.155277
     headage |   .0025086   .0028667     0.88   0.382      -.00311    .0081271
     eduhead |   .0083883   .0102168     0.82   0.412    -.0116363    .0284128
          ae |   .0115199   .0188305     0.61   0.541    -.0253872     .048427
       fhead |  -.2404342   .1166002    -2.06   0.039    -.4689663    -.011902
        rain |  -.0010461   .0008102    -1.29   0.197     -.002634    .0005419
      rain10 |   .0024185   .0008974     2.70   0.007     .0006598    .0041773
    stress10 |   .6507675   .1069308     6.09   0.000      .441187     .860348
        cv10 |  -.0271748    .015012    -1.81   0.070    -.0565978    .0022481
     Dtarmac |  -.0103837   .0017809    -5.83   0.000    -.0138742   -.0068932
   DKMsstall |   -.000832   .0067652    -0.12   0.902    -.0140915    .0124275
    avRTCsst |   .2813955   .6930832     0.41   0.685    -1.077023    1.639814
       _cons |   -2.59999   .9821629    -2.65   0.008    -4.524994   -.6749863
------------------------------------------------------------------------------
```

*** Model ii: Stage 4: SB quantity regression**

```
lnormal SSTmzkg $indep [pw=wei] if Afarm==1 & sell==1 & private==1 & lgmill==0

(Iteration log output omitted)

                                                  Number of obs     =      2,784
                                                  Wald chi2(13)     =     220.90
Log pseudolikelihood = -3876748.1                 Prob > chi2       =     0.0000

------------------------------------------------------------------------------
             |               Robust
     SSTmzkg |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      Bshare |   1.264162   .4400842     2.87   0.004     .4016126    2.126711
      Cshare |   1.875562   .5859757     3.20   0.001     .7270706    3.024053
     headage |   .0051101   .0021127     2.42   0.016     .0009692    .0092509
     eduhead |   .0658236   .0084291     7.81   0.000     .0493029    .0823443
          ae |   .0491194   .0143647     3.42   0.001     .0209651    .0772738
       fhead |  -.2479576   .0707986    -3.50   0.000    -.3867203   -.1091949
        rain |   .0013589   .0005138     2.64   0.008     .0003519    .0023658
      rain10 |  -.0027263   .0006498    -4.20   0.000    -.0039998   -.0014527
    stress10 |   .1911385   .0693454     2.76   0.006     .0552241    .3270529
        cv10 |  -.0437566   .0128536    -3.40   0.001    -.0689492    -.018564
     Dtarmac |   -.004437    .000922    -4.81   0.000    -.0062442   -.0026299
   DKMsstall |   .0042285    .005446     0.78   0.437    -.0064455    .0149025
    avRTCsst |   .0715705   .4801125     0.15   0.881    -.8694327    1.012574
       _cons |   6.548465   .6170118    10.61   0.000     5.339144    7.757786
-------------+----------------------------------------------------------------
    /lnsigma |    .100301   .0179108     5.60   0.000     .0651964    .1354056
-------------+----------------------------------------------------------------
       sigma |   1.105504   .0198005                      1.067369    1.145001
------------------------------------------------------------------------------
```

*** Model ii: Stage 4: LB quantity regression**

```
lnormal LSTmzkg $indep [pw=wei] if Afarm==1 & sell==1 & private==1 & lgmill==1

(Iteration log output omitted)

                                                  Number of obs     =        501
                                                  Wald chi2(13)     =     105.07
Log pseudolikelihood = -684011.54                 Prob > chi2       =     0.0000

------------------------------------------------------------------------------
             |               Robust
     LSTmzkg |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      Bshare |   -2.63701    1.00206    -2.63   0.008    -4.601011   -.6730094
      Cshare |   2.775902   1.353311     2.05   0.040     .1234608    5.428344
     headage |   .0052293   .0035928     1.46   0.146    -.0018125    .0122712
     eduhead |   .0321436   .0148353     2.17   0.030      .003067    .0612202
          ae |   .0752246   .0268778     2.80   0.005      .022545    .1279041
       fhead |  -.3822831   .1752134    -2.18   0.029    -.7256951   -.0388711
        rain |   .0002909   .0013022     0.22   0.823    -.0022613    .0028431
      rain10 |   -.003541   .0015567    -2.27   0.023    -.0065921   -.0004899
    stress10 |   .1665441   .1715382     0.97   0.332    -.1696646    .5027528
        cv10 |  -.0165189    .021103    -0.78   0.434      -.05788    .0248422
     Dtarmac |  -.0044536   .0050351    -0.88   0.376    -.0143222    .0054149
   DKMsstall |  -.0116588   .0271476    -0.43   0.668    -.0648671    .0415496
    avRTCsst |   1.778207    .690808     2.57   0.010     .4242478    3.132166
       _cons |   9.364135   1.635746     5.72   0.000     6.158132    12.57014
-------------+----------------------------------------------------------------
    /lnsigma |  -.1351897   .0368135    -3.67   0.000    -.2073429   -.0630366
-------------+----------------------------------------------------------------
       sigma |   .8735502   .0321584                      .8127409    .9389091
------------------------------------------------------------------------------
```

*** Model ii: Bootstrap the full model to predict all partial effects**

```
bootstrap       dPsell_dB=r(dPsell_dB)          ///
                        dPsell_dC=r(dPsell_dC)          ///
                        dPpriv_dB=r(dPpriv_dB)          ///
                        dPpriv_dC=r(dPpriv_dC)          ///
                        dPlgmil_dB=r(dPlgmil_dB)        ///
                        dPlgmil_dC=r(dPlgmil_dC)        ///
                        d1_B=r(d1_B)                    ///
                        d1_C=r(d1_C)                    ///
                        d2_B=r(d2_B)                    ///
                        d2_C=r(d2_C)                    ///
                        dQsst_dB=r(dQsst_dB)            ///
                        dQsst_dC=r(dQsst_dC)            ///
                        dQlst_dB=r(dQlst_dB)            ///
                        dQlst_dC=r(dQlst_dC)            ///
                        if Afarm==1, seed(39) reps(200): APEboot
(Iteration log output omitted)
------------------------------------------------------------------------------
             |   Observed   Bootstrap                         Normal-based
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   dPsell_dB |   .1187752   .0855724     1.39   0.165    -.0489436     .286494
   dPsell_dC |   .6740729   .1186631     5.68   0.000     .4414975    .9066482
   dPpriv_dB |   .2878547   .1447782     1.99   0.047     .0040947    .5716148
   dPpriv_dC |   .9828709   .1799188     5.46   0.000     .6302365    1.335505
  dPlgmil_dB |   .0630065   .1118349     0.56   0.573    -.1561859    .2821989
  dPlgmil_dC |   .5680254   .1555318     3.65   0.000     .2631887     .872862
        d1_B |    1109.34   405.8326     2.73   0.006     313.9226    1904.757
        d1_C |   1645.862   505.3796     3.26   0.001     655.3363    2636.388
        d2_B |  -7188.961   3020.024    -2.38   0.017     -13108.1   -1269.823
        d2_C |   7567.605   3776.654     2.00   0.045     165.4991    14969.71
    dQsst_dB |   323.5772   98.05586     3.30   0.001     131.3913    515.7632
    dQsst_dC |   740.1896   128.0301     5.78   0.000     489.2552     991.124
    dQlst_dB |  -84.52841    86.6844    -0.98   0.329    -254.4267     85.3699
    dQlst_dC |   661.4834   112.2411     5.89   0.000     441.4949    881.4719
------------------------------------------------------------------------------
```

**Model iii) 4-hurdle model w/hh; weather & transaction vars & prov*year**

```
*Model iii: Stage 1
global indep Bshare Cshare headage eduhead ae fhead rain rain10 stress10 cv10 Dtarmac DKMssta
avRTCsst
>       P1-P10_15

probit sell $indep [pw=wei] if Afarm==1

note: P10 omitted because of collinearity

(Iteration log output omitted)

Probit regression                                 Number of obs     =     15,082
                                                  Wald chi2(32)     =     825.31
                                                  Prob > chi2       =     0.0000
Log pseudolikelihood = -1763492.6                 Pseudo R2         =     0.0723
------------------------------------------------------------------------------
             |               Robust
        sell |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      Bshare |   1.770818   .2978188     5.95   0.000     1.187104    2.354532
      Cshare |   .4137052   .3612497     1.15   0.252    -.2943311    1.121742
     headage |  -.0009236    .000942    -0.98   0.327    -.0027698    .0009227
     eduhead |   .0390469   .0041719     9.36   0.000       .03087    .0472237
          ae |   .0434756   .0069143     6.29   0.000     .0299239    .0570274
       fhead |  -.1121617   .0352737    -3.18   0.001    -.1812969   -.0430265
        rain |   .0000657   .0003706     0.18   0.859    -.0006608    .0007922
      rain10 |  -.0008625   .0005153    -1.67   0.094    -.0018724    .0001475
    stress10 |  -.2713623   .0609301    -4.45   0.000     -.390783   -.1519415
        cv10 |   .0318038    .009138     3.48   0.001     .0138936     .049714
     Dtarmac |  -.0016003   .0005728    -2.79   0.005    -.0027229   -.0004777
   DKMsstall |  -.0139545   .0027009    -5.17   0.000    -.0192482   -.0086608
    avRTCsst |  -.3914198   .3096082    -1.26   0.206    -.9982407    .2154011
          P1 |   .9514549   .0978194     9.73   0.000     .7597323    1.143177
          P2 |   .8978784   .1057722     8.49   0.000     .6905686    1.105188
          P3 |   .7240375   .0938686     7.71   0.000     .5400583    .9080166
          P4 |   .5733097   .1001922     5.72   0.000     .3769366    .7696827
          P5 |   .7092475   .1259286     5.63   0.000     .4624319    .9560631
          P6 |   1.201088    .106266    11.30   0.000     .9928108    1.409366
          P7 |   .7766405   .0962572     8.07   0.000     .5879798    .9653011
          P8 |   1.048896   .1007835    10.41   0.000     .8513639    1.246428
          P9 |   .3353208   .1056675     3.17   0.002     .1282163    .5424253
         P10 |          0  (omitted)
       P1_15 |   .1078049    .103178     1.04   0.296    -.0944202    .3100299
       P2_15 |   .2920417   .1007265     2.90   0.004     .0946213    .4894621
       P3_15 |   .0301694   .0570317     0.53   0.597    -.0816108    .1419495
       P4_15 |   .1183228   .0875267     1.35   0.176    -.0532264    .2898719
       P5_15 |  -.0120908   .1364341    -0.09   0.929    -.2794967    .2553152
       P6_15 |    .023377   .0932462     0.25   0.802    -.1593823    .2061363
       P7_15 |   .0641668    .086432     0.74   0.458    -.1052369    .2335704
       P8_15 |   .1532467   .1038141     1.48   0.140    -.0502252    .3567185
       P9_15 |  -.0093202   .0956783    -0.10   0.922    -.1968462    .1782058
      P10_15 |   .0642458   .1071378     0.60   0.549    -.1457404     .274232
       _cons |  -.9442053   .4112425    -2.30   0.022    -1.750226   -.1381848
------------------------------------------------------------------------------
```

***Model iii: Stage 2**

```
probit private $indep [pw=wei] if Afarm==1 & sell==1

note: P10 omitted because of collinearity

(Iteration log output omitted)

Probit regression                               Number of obs     =      7,212
                                                Wald chi2(32)     =     531.65
                                                Prob > chi2       =     0.0000
Log pseudolikelihood =  -738551.4               Pseudo R2         =     0.1014
------------------------------------------------------------------------------
             |               Robust
     private |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      Bshare |   .3391737   .4336723     0.78   0.434    -.5108084    1.189156
      Cshare |    .063095   .5083774     0.12   0.901    -.9333064    1.059496
     headage |  -.0032746   .0014839    -2.21   0.027     -.006183   -.0003662
     eduhead |  -.0306676   .0060151    -5.10   0.000     -.042457   -.0188782
          ae |  -.0498851   .0097673    -5.11   0.000    -.0690288   -.0307415
       fhead |  -.0131391   .0568731    -0.23   0.817    -.1246083    .0983302
        rain |    .000354   .0005603     0.63   0.528    -.0007443    .0014523
      rain10 |  -.0011767   .0008017    -1.47   0.142     -.002748    .0003947
    stress10 |  -.1650844   .0968182    -1.71   0.088    -.3548447    .0246758
        cv10 |   .0083579    .013798     0.61   0.545    -.0186857    .0354015
     Dtarmac |   .0042549   .0009396     4.53   0.000     .0024134    .0060965
   DKMsstall |   .0087456   .0051796     1.69   0.091    -.0014061    .0188974
    avRTCsst |   .1836954   .6317759     0.29   0.771    -1.054563    1.421953
          P1 |  -.2419539   .1669437    -1.45   0.147    -.5691576    .0852499
          P2 |  -.2023543   .1790648    -1.13   0.258    -.5533149    .1486063
          P3 |  -.4893025   .1700747    -2.88   0.004    -.8226429   -.1559621
          P4 |   -.284499   .1837213    -1.55   0.121    -.6445861     .075588
          P5 |  -.2120182   .2104547    -1.01   0.314    -.6245017    .2004654
          P6 |   -.927076   .1844407    -5.03   0.000    -1.288573   -.5655789
          P7 |  -.8614558    .172647    -4.99   0.000    -1.199838    -.523074
          P8 |  -.3156483   .1755709    -1.80   0.072    -.6597609    .0284643
          P9 |  -.2147469   .1855908    -1.16   0.247    -.5784982    .1490043
         P10 |          0  (omitted)
       P1_15 |    .729803   .1297979     5.62   0.000     .4754038    .9842022
       P2_15 |   .8466232   .1384606     6.11   0.000     .5752454    1.118001
       P3_15 |   .5659725   .0883861     6.40   0.000     .3927389    .7392062
       P4_15 |  -.0272414   .1430538    -0.19   0.849    -.3076217    .2531388
       P5_15 |   .9527747   .1996913     4.77   0.000     .5613869    1.344162
       P6_15 |   .0851321   .1416898     0.60   0.548    -.1925749    .3628391
       P7_15 |   .2098589   .1290212     1.63   0.104    -.0430181    .4627358
       P8_15 |   .0752885   .1334817     0.56   0.573    -.1863307    .3369078
       P9_15 |   .6990443   .1527947     4.58   0.000     .3995722    .9985165
      P10_15 |   .5363529   .2267397     2.37   0.018     .0919512    .9807546
       _cons |   1.325959   .6702937     1.98   0.048     .0122077    2.639711
------------------------------------------------------------------------------
```

***Model iii: Stage 3**

```
probit lgmill $indep [pw=wei] if Afarm==1 & sell==1 & private==1

note: P10_15 != 0 predicts failure perfectly
      P10_15 dropped and 86 obs not used
```

*(Iteration log output omitted)*

```
Probit regression                               Number of obs     =      3,199
                                                Wald chi2(31)     =     207.84
                                                Prob > chi2       =     0.0000
Log pseudolikelihood = -206556.63               Pseudo R2         =     0.1187
------------------------------------------------------------------------------
             |               Robust
      lgmill |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      Bshare |   1.840135   .8888924     2.07   0.038     .0979384    3.582333
      Cshare |  -.3573716   .8467085    -0.42   0.673     -2.01689    1.302147
     headage |   .0014313   .0030219     0.47   0.636    -.0044916    .0073541
     eduhead |   .0128308   .0110085     1.17   0.244    -.0087454     .034407
          ae |    .010939   .0191419     0.57   0.568    -.0265784    .0484563
       fhead |  -.1962838   .1195762    -1.64   0.101    -.4306489    .0380813
        rain |   .0004436   .0012139     0.37   0.715    -.0019356    .0028228
      rain10 |    .001624   .0017611     0.92   0.356    -.0018277    .0050756
    stress10 |   .0334717    .170066     0.20   0.844    -.2998516    .3667949
        cv10 |   .0638824     .02474     2.58   0.010     .0153928    .1123719
     Dtarmac |  -.0053077   .0019275    -2.75   0.006    -.0090856   -.0015298
   DKMsstall |  -.0176586   .0122139    -1.45   0.148    -.0415974    .0062803
    avRTCsst |   .9211479   .8236465     1.12   0.263    -.6931695    2.535465
          P1 |    .573616   .3709047     1.55   0.122    -.1533438    1.300576
          P2 |   .4552621   .4109073     1.11   0.268    -.3501014    1.260626
          P3 |    .244409   .3866929     0.63   0.527    -.5134951    1.002313
          P4 |  -.0636886   .4995575    -0.13   0.899    -1.042803    .9154261
          P5 |   .1362859   .4348728     0.31   0.754    -.7160491    .9886209
          P6 |   .7897911   .4190466     1.88   0.059    -.0315252    1.611107
          P7 |     .19854    .479301     0.41   0.679    -.7408727    1.137953
          P8 |  -.3857082   .4865325    -0.79   0.428    -1.339294     .567878
          P9 |  -.0483236    .389498    -0.12   0.901    -.8117256    .7150785
         P10 |          0  (omitted)
       P1_15 |   .3601921   .2025391     1.78   0.075    -.0367772    .7571615
       P2_15 |   .4281681    .219083     1.95   0.051    -.0012266    .8575629
       P3_15 |   1.049557   .1697242     6.18   0.000     .7169039     1.38221
       P4_15 |   .1393167   .4018561     0.35   0.729    -.6483067    .9269401
       P5_15 |   .5961957   .3336745     1.79   0.074    -.0577943    1.250186
       P6_15 |  -.2859392   .3125634    -0.91   0.360    -.8985522    .3266738
       P7_15 |  -.0398977   .3346492    -0.12   0.905     -.695798    .6160026
       P8_15 |   .6046506    .479597     1.26   0.207    -.3353423    1.544643
       P9_15 |  -.1116914   .2386127    -0.47   0.640    -.5793637     .355981
      P10_15 |          0  (omitted)
       _cons |  -4.323404   1.433405    -3.02   0.003    -7.132827   -1.513981
------------------------------------------------------------------------------
```

***Model iii: Stage 4 (SB quantity)**

```
lnormal SSTmzkg $indep [pw=wei] if Afarm==1 & sell==1 & private==1 & lgmill==0

note: P10 omitted because of collinearity

(Iteration log output omitted)

                                                  Number of obs     =      2,784
                                                  Wald chi2(32)     =     403.54
Log pseudolikelihood = -3930772.3                 Prob > chi2       =     0.0000


                                                  Number of obs     =      2,784
                                                  Wald chi2(32)     =     402.38
Log pseudolikelihood = -3856114.9                 Prob > chi2       =     0.0000
------------------------------------------------------------------------------
             |               Robust
     SSTmzkg |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      Bshare |   1.506141   .5765214     2.61   0.009     .3761799    2.636102
      Cshare |  -.2931995   .6285501    -0.47   0.641    -1.525135     .938736
     headage |   .0041863   .0020145     2.08   0.038     .0002379    .0081346
     eduhead |   .0674367   .0081737     8.25   0.000     .0514166    .0834568
          ae |   .0409835     .01358     3.02   0.003     .0143673    .0675997
       fhead |  -.2553954   .0669649    -3.81   0.000    -.3866442   -.1241466
        rain |    .000726   .0006901     1.05   0.293    -.0006265    .0020786
      rain10 |  -.0030444    .001063    -2.86   0.004    -.0051279   -.0009608
    stress10 |   .0853042   .1163871     0.73   0.464    -.1428104    .3134187
        cv10 |  -.0270421   .0196317    -1.38   0.168    -.0655196    .0114354
     Dtarmac |   -.002148   .0011046    -1.94   0.052    -.0043129     .000017
   DKMsstall |  -.0075323   .0041455    -1.82   0.069    -.0156573    .0005927
    avRTCsst |   .5571526   .5025547     1.11   0.268    -.4278365    1.542142
          P1 |   .4919802   .1710339     2.88   0.004       .15676    .8272004
          P2 |   .6929825   .1721765     4.02   0.000     .3555227    1.030442
          P3 |    .112868   .1688339     0.67   0.504    -.2180403    .4437763
          P4 |  -.0044826   .1916666    -0.02   0.981    -.3801422    .3711771
          P5 |   .0086829   .2371242     0.04   0.971     -.456072    .4734378
          P6 |   .0795053   .2247281     0.35   0.724    -.3609538    .5199643
          P7 |   .0520018   .1913315     0.27   0.786    -.3230011    .4270047
          P8 |   .8515766   .1684587     5.06   0.000     .5214036    1.181749
          P9 |   .3181969    .195199     1.63   0.103    -.0643861    .7007799
         P10 |          0  (omitted)
       P1_15 |   .4989182   .1826094     2.73   0.006     .1410104    .8568261
       P2_15 |   .0223461   .1561702     0.14   0.886    -.2837418     .328434
       P3_15 |    .557644   .1364947     4.09   0.000     .2901194    .8251686
       P4_15 |   .0679611   .1828736     0.37   0.710    -.2904645    .4263866
       P5_15 |   .6958399   .2565487     2.71   0.007     .1930138    1.198666
       P6_15 |   .2605854   .2317585     1.12   0.261    -.1936528    .7148236
       P7_15 |   .2495554   .1640618     1.52   0.128    -.0719998    .5711105
       P8_15 |   .2002738   .1438798     1.39   0.164    -.0817253     .482273
       P9_15 |   .2186933   .1686255     1.30   0.195    -.1118066    .5491932
      P10_15 |    .045042    .169302     0.27   0.790    -.2867838    .3768678
       _cons |   6.849946   .9165641     7.47   0.000     5.053514    8.646379
-------------+----------------------------------------------------------------
    /lnsigma |   .0600705   .0175088     3.43   0.001      .025754    .0943871
-------------+----------------------------------------------------------------
       sigma |   1.061911   .0185928                      1.026088    1.098985
------------------------------------------------------------------------------
```

***Model iii: Stage 4 (LB quantity)**

```
lnormal mzkg $indep [pw=wei] if Afarm==1 & sell==1 & private==1 & lgmill==1

note: P10 omitted because of collinearity
note: P10_15 omitted because of collinearity

(Iteration log output omitted)
                                                  Number of obs     =        501
                                                  Wald chi2(31)     =     205.11
Log pseudolikelihood = -680566.23                 Prob > chi2       =     0.0000

------------------------------------------------------------------------------
             |               Robust
     LSTmzkg |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      Bshare |  -2.452476   1.270391    -1.93   0.054    -4.942397    .0374442
      Cshare |    1.53958    1.46437     1.05   0.293    -1.330532    4.409693
     headage |    .005777   .0031335     1.84   0.065    -.0003646    .0119186
     eduhead |    .026693   .0149232     1.79   0.074     -.002556     .055942
          ae |   .0692426   .0250668     2.76   0.006     .0201125    .1183727
       fhead |  -.3683656   .1622911    -2.27   0.023    -.6864504   -.0502809
        rain |  -.0013771   .0018064    -0.76   0.446    -.0049177    .0021634
      rain10 |  -.0025846   .0032039    -0.81   0.420     -.008864    .0036949
    stress10 |  -.0213387   .2271241    -0.09   0.925    -.4664937    .4238164
        cv10 |  -.0205006   .0366003    -0.56   0.575    -.0922358    .0512346
     Dtarmac |  -.0016185   .0048665    -0.33   0.739    -.0111566    .0079195
   DKMsstall |  -.0355335   .0344272    -1.03   0.302    -.1030096    .0319426
    avRTCsst |   .7817651   .8369411     0.93   0.350    -.8586093    2.422139
          P1 |  -.4419415   .5026771    -0.88   0.379    -1.427171    .5432875
          P2 |  -.3709572   .4498091    -0.82   0.410    -1.252567    .5106525
          P3 |  -.3338944   .5187162    -0.64   0.520     -1.35056    .6827708
          P4 |   -.821912   .6596907    -1.25   0.213    -2.114882     .471058
          P5 |  -.6782222   .6893034    -0.98   0.325    -2.029232    .6727876
          P6 |  -.6347421   .4807134    -1.32   0.187    -1.576923    .3074388
          P7 |  -1.395705   .6978476    -2.00   0.045    -2.763461   -.0279486
          P8 |  -.7363417   .5384023    -1.37   0.171    -1.791591    .3189075
          P9 |  -.3909438   .6450999    -0.61   0.545    -1.655316    .8734288
         P10 |          0  (omitted)
       P1_15 |   .2570228   .2679214     0.96   0.337    -.2680936    .7821391
       P2_15 |   .3821709   .3136863     1.22   0.223    -.2326428    .9969847
       P3_15 |   .0067525   .2985856     0.02   0.982    -.5784646    .5919695
       P4_15 |    .079191     .64316     0.12   0.902    -1.181379    1.339761
       P5_15 |   .7727631   .5347248     1.45   0.148    -.2752782    1.820804
       P6_15 |   .3323543   .3902681     0.85   0.394     -.432557    1.097266
       P7_15 |   1.094321   .5805568     1.88   0.059    -.0435495    2.232191
       P8_15 |  -.3899027   .4558396    -0.86   0.392    -1.283332    .5035266
       P9_15 |   .7139772   .3812166     1.87   0.061    -.0331936    1.461148
      P10_15 |          0  (omitted)
       _cons |    10.5294   2.187809     4.81   0.000     6.241373    14.81743
-------------+----------------------------------------------------------------
    /lnsigma |  -.1775318   .0378063    -4.70   0.000    -.2516306   -.1034329
-------------+----------------------------------------------------------------
       sigma |   .8373344   .0316565                      .7775319    .9017366
------------------------------------------------------------------------------
```

*** Model iii: Bootstrap the full model to predict all partial effects**

```
bootstrap       dPsell_dB=r(dPsell_dB)          ///
                dPsell_dC=r(dPsell_dC)          ///
                dPpriv_dB=r(dPpriv_dB)          ///
                dPpriv_dC=r(dPpriv_dC)          ///
                dPlgmil_dB=r(dPlgmil_dB)        ///
                dPlgmil_dC=r(dPlgmil_dC)        ///
                d1_B=r(d1_B)                    ///
                d1_C=r(d1_C)                    ///
                d2_B=r(d2_B)                    ///
                d2_C=r(d2_C)                    ///
                dQsst_dB=r(dQsst_dB)            ///
                dQsst_dC=r(dQsst_dC)            ///
                dQlst_dB=r(dQlst_dB)            ///
                dQlst_dC=r(dQlst_dC)            ///
                if Afarm==1, seed(40) reps(200): APEboot
```

*(Iteration log output omitted)*

```
Bootstrap results                               Number of obs     =     15,087
                                                Replications      =        158

------------------------------------------------------------------------------
             |   Observed   Bootstrap                         Normal-based
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   dPsell_dB |   .6399975   .1059155     6.04   0.000     .4324069    .8475882
   dPsell_dC |   .1495187   .1242391     1.20   0.229    -.0939855    .3930228
   dPpriv_dB |   .1205706   .1541427     0.78   0.434    -.1815435    .4226846
   dPpriv_dC |   .0224292   .2016299     0.11   0.911    -.3727581    .4176165
  dPlgmil_dB |   .3643602   .1796013     2.03   0.042     .0123481    .7163724
  dPlgmil_dC |  -.0707622   .1761094    -0.40   0.688    -.4159303    .2744059
        d1_B |   1300.219   566.3832     2.30   0.022     190.1282     2410.31
        d1_C |  -253.1128   549.0091    -0.46   0.645    -1329.151    822.9252
        d2_B |  -6709.542   3716.777    -1.81   0.071    -13994.29    575.2061
        d2_C |    4212.02   4764.211     0.88   0.377    -5125.661     13549.7
    dQsst_dB |   424.4937   132.0257     3.22   0.001     165.7281    683.2593
    dQsst_dC |   28.82568   129.9819     0.22   0.824    -225.9342    283.5856
    dQlst_dB |   112.6033   119.1117     0.95   0.344    -120.8514    346.0579
    dQlst_dC |   95.26385   147.0381     0.65   0.517    -192.9255    383.4532
------------------------------------------------------------------------------
Note: One or more parameters could not be estimated in 42 bootstrap replicates;
      standard-error estimates include only complete replications.
```