Transaction Costs and Organic Marketing: Evidence from U.S. Organic Produce Farmers

Bo MacInnis
Dept. of Agricultural and Resource Economics
University of California, Berkeley
macinnis@are.berkeley.edu

Abstract: We develop a conceptual framework that integrates quality of output and transaction costs in the choice of marketing channels; based upon which, we estimate a reduced-form Tobit model and a semi-reduced logit model with a farm-level cross-sectional dataset to measure the effects of transaction costs in farmer's ability to make sales to indirect markets (retailers and wholesalers). We find strong empirical evidence that existing organic retail and wholesale markets impose considerable barriers to entry to individual organic farmers, and furthermore, the effects of transaction costs are asymmetric between the two types of farmers, those who transitioned from conventional farming and those who did not. Those who did are overall favored, and those who did not are constrained by more types of transaction costs and are constrained more severely than those who did. We argue that an effect policy should target to the least favored farmers by encouraging or mandating distributors and retailers have a more transparent and objective process in selecting organic suppliers, such that all farmers would have an equal opportunity to be successful in selling to indirect markets.

* I thank Gordon Rausser, Guido Imbens and Larry Karp for their encouragement, patience and insightful suggestions. All remaining errors are my own. This is a working draft of an ongoing research. Your suggestions are welcome.

Introduction

Relying on biologically and ecologically based practices, organic farming virtually excludes the use of synthetic chemicals such as synthetic fertilizers and pesticides of all kinds, antibiotics, and hormones in crop production; and it prohibits the use of antibiotics and hormones in livestock production[i]. A farmer must be inspected and certified by an accredited certification agency after farming under the organic farming standards for at least three years. Only after being certified organic can farmers market their agricultural products as “organic”.

Organic farming is one of the fastest growing segments of U.S. agriculture. Certified organic farmland for corn, soybeans, and livestock sectors doubled from 1992 to 1997, and doubled again between 1997 and 2001. There are 2.3 million acres of cropland and pasture dedicated to organic production in 2001. (Greene 2001)

Consumer demand for organic food has been rising steadily. The organic food market is among the fastest growing categories in U.S. food industry. While the food industry has grown 1% or less annually in the last decade, retail sales of organic food products have grown more than 20% over the same period. Organic sales grew from $1 billion in 1990 to $7.8 billion in 2000. Organic food is available in 73% of mainstream grocery stores, and more than half of the total organic sales take place in traditional retailers[ii].

Irrespective of the high growth rates, organic production remains an inconsequentially small fraction of U.S. agriculture: 0.3 percent of all farmland is certified organic, and organic food sales represents 1.3 percent of total food expenditure[iii]. One would ask why organic farming is still so small. USDA identified obstacles of adopting organic farming as follows: high managerial costs, risks of shifting to a new way of farming, limited awareness of organic farming systems, lack of organic marketing and infrastructure, and inability to capture price premiums[iv]. These obstacles are confirmed in a recent survey by Organic Farming Research Foundation, where the most severe barriers to transitioning indicated were lack of information and experience in organic production and an inability to identify markets for organic products (Waltz).

Organic produce do not sell themselves; farmers must sell them. Selling incurs transaction costs. Transaction costs of selling in organic markets may be significantly high because the organic markets are thin, and necessary institutions and infrastructure are not yet fully-developed[v]. If the barriers to entry to organic markets are prohibitively high to many organic farmers, expansion of organic production and markets would be severely constrained.

There is a scarce body of literature on organic marketing[vi]. This paper intends to apply the transaction costs economics framework (Williams and Hobbs) to the choice of marketing channels for the organic produce farmers in the United States. We focus on identifying and measuring the barriers to entry to retail and wholesale markets, because penetrating these markets is essential to sustain and expedite the growth of organic farming. We ask two questions: (i) what kinds of transaction costs do the existing retail and wholesale markets impose to organic farmers? which transaction costs are more significant than others? (ii) do these transaction costs affect all organic farmers in the same way?

The rest of the paper is organized as follows. Section 1 provides a brief literature review. Section 2 describes the data and preliminary data analysis. Section 3 develops a theoretical model that motivates the econometric specifications. Section 4 is the econometric analysis. Section 5 concludes.

1. Literature Review

Market transactions do not occur in a frictionless environment. Transaction costs are economic equivalent to frictions in physical systems. Transaction costs are often categorized into ex ante and ex post transaction costs. Both types of transaction costs are interdependent, and their relative importance depends on the nature and frequency of transactions (Williamson).

1.1 Transaction Costs and Agricultural Marketing

Transaction costs are not available on financial records, and inherently difficult to measure or quantify. There are a number of empirical studies on the effects of transaction costs on agricultural marketing despite the measurement difficulty.

The effects of transaction costs in marketing agricultural products are well-studied in transition and developing economies where markets are thin and fledging, and necessary infrastructure is missing or embryonic. Goetz, Omamo, and Key, Sadoulet and de Janvary use the agricultural household model and investigate the effect of transaction costs on the joint decisions of market participation and supply responses. Hobbs, Bailey and Hunnicutt, and Ferto and Szabo analyze the role of transaction costs in agricultural market selection in both transition economy and developed economy.

Hobbs is an influential work in applying the transaction cost economics framework to the choice of marketing channels in agricultural products. Hobbs identifies three types of transaction costs in agricultural marketing: information costs, negotiation and bargaining costs, and monitoring and enforcement costs. A form of ex ante transaction costs, information costs are the costs of identifying markets and trading partners, and costs of obtaining price and product information. Negotiation costs are the costs of physically carrying out the transaction, including the costs of physically negotiating, bargaining and formally drawing up the terms of exchange. A form of ex post transaction costs, monitoring and enforcement costs are the costs of ensuring that the trading partners follow the terms of the transaction, such as quality standards or payment arrangements.

Recognizing the difficulty of obtaining transaction costs data, Hobbs demonstrates a method for measuring the influence of transaction costs on the choice of cattle marketing between live-ring auction and direct-to-packers. She uses a survey data of UK cattle farmers and a two-limit Tobit model to estimate the relative importance of various transaction costs and farm characteristics on channel selection. She found that information costs are not statistically significant, but negotiation and monitoring costs are significant in the UK cattle auction market.

Following Hobbs, Bailey and Hunnicutt measures the importance of transaction costs in market selection by Utah commercial feeder cattle producers. They use a seemingly unrelated regression model and find that implicit transaction costs, such as relationships and experience are important. Ferto and Szabo applies Hobbs's methodology to the choice of supply channels in Hungarian fruit and vegetable sector using a multinomial logit model. They found that the choice of selling to wholesale market is strongly and negatively affected by information costs and farmer's age, and negatively by the bargaining power and monitoring costs.

1.2 Organic Agricultural Policy

There are several studies on organic agriculture policy. Lampkin and Padel attributes the high conversion levels in the European Union countries to government's intervention such as developing consumer education initiatives and providing conversion subsidies.

Pietola and Lansink uses an optimal stopping model and estimates the effect of conversion subsidy on the adoption of organic farming in Finland between 1994 and 1997. They find that decreasing output prices and increasing direct subsidies trigger the switch to organic farming, furthermore, the switch is more likely for farmers that have large land areas and low yields, and the switch is less likely for farms with intensive livestock production and labor-intensive production.

Lohr and Salomonsson uses an random utility model to compare farmers in Sweden in 1990 who converted before and after the subsidy. They find that greater livestock diversity and more sales outlets are significant conversation factors without subsidies. Their results suggest that a marketing and technical information infrastructure designed to support conventional agriculture restrict the potential effect of a conversion subsidy in the United States.

Intervention by USDA on organic agriculture has focused primarily on market facilitation, such as establishing federal standards and labels including the release of Nation Organic Program in 2002, and adding several initiatives to assist organic farmers in the 2002 Farm Act. There has never been conversion subsidy at the federal level, though several states – Minnesota and Iowa in particular – have begun subsidizing the adoption of organic farming systems, and the effects of these subsidy programs are yet to evaluated. (Greene 2003)

2. The Data

The dataset we use in this study is the 1997 nationwide organic farmers survey conducted by Organic Farming Research Foundation[vii]. The Organic Farming Research Foundation is a non-profit whose mission is to sponsor research related to organic farming practices, to disseminate research results to organic farmers and to growers interested in adopting organic production systems, and to educate the public and decision-makers about organic farming issues. Founded in 1989, Organic Farming Research Foundation conducts nationwide surveys on organic farmers every four years since 1990. Their surveys are recognized to be the most comprehensive micro level data on organic farming. The 1997 survey is the third national survey and is the most recent survey available to research at the time of writing. However, the 2001 survey is complete and not made available to public yet.

2.1 The Survey

The survey is a cross-sectional farm-level data on production, marketing and demographics. The survey was sent to 4,638 certified organic farmers from fifty-five organic certification organizations[viii]. 1,192 surveys were returned from organic farmers in forty-five states. Samples contain farmers who grow one or more types of the three major agricultural products: fresh produce and herbs, field crops, and livestock animals. The samples we use are those for-profit farms that grow produce only. After discarding samples that produced livestock animals or field crops, we have 396 usable samples.

The survey contain six sections on organic farming research priority, information resources, organic products grown, organic marketing, organic production management concerns and strategies, organic production and marketing constraints and challenges, organic certification, and farm management and demographics. We investigate the effects of transaction costs in marketing organic produce, and we find the data on organic production management concerns and strategies irrelevant to our study, because this data is about the detailed production process and we assume production process and marketing decisions are separable. We also find the data on organic certification irrelevant, because certification must occur prior to selling to organic markets, and all samples have already been certified.

We use a subset of the survey, mainly from the data of four sections: organic products grown, organic marketing, organic production and marketing constraints and challenges, and farm management and demographics. The organic marketing data provides how farmers allocate their output to a number of channels, which we aggregate them into two board categories: direct and indirect channels. This is the variable we intend to explain and predict.

The organic production and marketing constraints and challenges data provides detailed transaction costs in marketing the outputs from the farmer's perspective. Following Hobbs, we categorize transaction costs data into information costs, negotiation costs, monitoring costs, and market characteristics. Definition and measurement of the variables are described in Table 1.

We use the following variables as proxies for information costs: costs of finding organic markets, of obtaining access to existing markets and of searching for best prices. Variables that can be used for proxies for negotiation or monitoring costs are limited in this survey. We use the distance between producer and market or delivery point as the proxy for negotiation cost, and failure of buyers to honor commitment and reliable or prompt payment as two proxies for monitoring costs.

As a form of ex ante transaction costs, market environment data measures the level of opportunity and frictions to transact (Hobbs). The market characteristics we consider include lack of acceptance of certification documents in certain markets, oversupply of legitimate organic products in existing markets, and lack of consumer understanding about organic food. In addition, we construct an index of total available markets for direct and indirect channels. We use the number of farmers' markets in each state as an index for direct market infrastructure[ix]. We use the number of organic retailers, processors and manufacturers, and wholesalers in each state as an index for indirect market infrastructure[x]. Since indirect markets are often accessible across states, we take account for the effect of cross-state spillover, which is calculated as half of the weighted average of the indexes of the adjacent states.

We also use a set of socioeconomic and farm characteristics data, and its definitions and measurement are described in Table 2. There are four reasons to include these variables in the analysis. First, we wish to investigate the heterogeneity of preferences and risk attitudes by individual characteristics such as gender, age, education, and experience of the farmer, as well as farm characteristics such as business structure and land size. Second, we allow the heterogeneity of quality distribution of the output, which in turn allows for the heterogeneous transaction costs, by using individual characteristics as proxies. Third, there is non-response (response rate is 26 percent), which may be correlated with individual differences as well as channel allocation decisions. Covariates can be used to adjust for these differences[xi]. Fourth, the inclusion of control variables can improve the precision of the estimates.

2.2 Summary Statistics

Among the 1,192 observations, we discard samples that are not for-profit farms as well as that produced livestock animals or field crops with and without growing fresh produce. This leaves us 1,152 samples. We also discard samples with missing data to some of the questions in the survey. This leaves us 360 usable samples. By simply discarding the observations with missing data, we may have introduced biases or at least lost some precision. Alternative approaches to item non-response without losing efficiency or introducing biases would be worthwhile pursuing in the future.

There are two types of organic farmers: the transitioners are the farmers who transitioned to organic farming from conventional farming, and the beginners are the farmers who started organic farming without prior conventional farming experience. A somewhat surprising and definitely interesting thing is the beginners constitute a larger fraction of the sample than the transitioners – 59 percent in the 1,152 for-profit farm sample (686 beginners and 476 transitioners) and 69 percent in the 360 usable samples for the analysis (249 beginners and 111 transitioners).

We cannot rule out the possibility that both unit non-response error and item non-response error are correlated with whether the sample is a transitioner or beginner. For examples, are the beginners more likely to respond to the survey than the transitioners? in addition, are they more likely to respond more completely to the questions in the survey than the transitioners? Our treatment on both kinds of non-response errors are rudimentary. As discussed earlier, we use covariates to help adjust for the non-response errors assuming that non-response errors are correlated with individual characteristics. We are interested in investigating the non-response errors further in our future work.

Whether or not the beginners are over-sampled because of the non-response error, it is interesting that organic farming has attracted many new entrants, those who did not farm conventionally in the past. The transitoners and the beginners differ in a number of substantial ways – reasons for adopting organic farming, experience in agricultural production, and experience in agricultural marketing. For this reason, we investigate the difference between these two sub-groups of organic farmers in more details throughout the paper.

Table 3 present the summary statistics of all the variables used in the analysis. For each variable the mean and standard deviation for the entire sample are listed in the first two columns. We also present the means separately for the transitioners and beginners in next two columns. The last column is the t-statistics for the null hypotheses that the averages for the transitioners and beginners subsamples are identical.

The transitioners sell significantly greater proportion of their output to indirect markets than the beginners (80 % versus 51 %, *t*-stat is 2.36). Three transaction cost variables differ significantly between the transitioners and the beginners: obtaining market access (2.56 versus 2.27, t-stat is 2.64), lack of consumer understanding (2.93 versus 3.06, *t*-stat is 4.46), and reliable payment (2.12 versus 2.07, *t*-stat is 2.35). The transitioners perceive more difficulty in obtaining market access than the beginners. Lack of consumer understanding and reliable payment are more severe constraints to the transitioners than to the beginners[xii]. In addition, the transitioners are located in states where there are greater level of organic marketing infrastructure, both for direct markets and indirect markets, than the beginners, though the differences are not statistically significant.

Except for age, most socioeconomic characteristics differ between the transitioners and beginners subsamples, though not all differences are statistically significant. For the production characteristics, the beginners subsample, on average, grow more varieties of produce (7.2 versus 3.3) and make more value-added products (0.63 versus 0.30) than the transitioners subsample, though neither is statistically significant.

For the farm characteristics, the transitioners come from more formal business structures, and their farms have been certified for shorter duration (4.62 years versus 5.92 years) than the beginners. The beginners have larger size of land dedicated to organic farming than the transitioners (128 acres versus 98 acres), but the total acres of farming land are very similar, which suggests that the transitioners are more likely to have a mixed operation where both conventional and organic farming are practiced. In addition, the transitioners own considerably more land than the beginners (135 acres versus 45 acres). None of these means differ significantly, however.

For the farmer characteristics, the beginners have more experience in organic farming (10.33 years versus 8.55 years) but less experience in farming including organic and conventional farming (14.49 years versus 15.63 years). Neither measure of farming experience differ significantly. Two variables do differ significantly between the transitioners and the beginners: education (4.88 versus 4.73, *t*-stat is 2.82), and gender (1.81 versus 1.78, *t*-stat is 17.61). The transitioners are more likely to complete college and more likely to be male than the beginners.

2.3 Preliminary Data Analysis

Based on the summary statistics, we perform two tasks in the preliminary data analysis: to see whether transaction costs are significant barriers to entry to indirect markets, and to see whether transaction costs impose same levels of barriers of entry to the transitioned organic farmers (the transitioners) and the new entrants organic farmers (the beginners).

We regress the proportion sold to indirect markets to all the transaction costs variables using a two-limit tobit model on three sets of samples separately: the whole sample, the transitioners subsample and the beginners subsample. Table 4 presents the estimation results. First, a number of transaction costs variables are statistically significant for the entire sample as well as for the two subsamples, such as finding markets, reliable payment, lack of consumer understanding, and direct and indirect marketing network. Second, several transaction costs variables exhibit similar effects on the two subsamples, such as reliable payment, lack of consumer understanding and direct and indirect marketing network. These three variables have similar coefficients as well as the high level of statistical significance. Third, three variables – obtaining access, distance and over-supply, have larger magnitudes of coefficients and are significant in the whole sample, but have smaller yet similar magnitudes of coefficients and are not significant in either subsample. Fourth, two variables – finding markets and lack of acceptance of certification, have different magnitudes of estimates and statistical significance levels for the two subsamples.

We use a Hausman test for the null hypothesis that the coefficients in two subsample estimation are identical; the test statistics χ^2 (11) = 15.69 and p-value is 0.1529, and we reject the null. In addition, we reject the null hypothesis that all coefficients in the entire sample and the transitioners subsample are identical with χ^2 (11) = 8.42 and p-value 0.675; we fail to reject the null hypothesis that the coefficients in the entire sample and the beginners subsample are identical with χ^2 (11) = 53.04 and p-value 0.000. This suggests that not all transaction costs affect the two subamples in the same way. Investigating the heterogeneous effects of transaction costs has direct policy implication whereby it may be more efficient to reduce one kind of transaction for one group of organic farmers and to reduce a different kind of transaction costs for another group of organic farmers.

3. Theoretical Model

Farmers make production and marketing decisions. The former concerns the portfolio of crops, and land allocation and input uses for each crop; the latter concerns finding and obtaining access to the markets, and allocating and selling output to relevant marketing channels. Both decisions are inter-related and should be simultaneously determined. In this study, we model farmers' marketing decisions assuming the production decisions are pre-determined, as we are constrained by the data.

We focus on the organic produce sector for two reasons: fresh vegetable and fruits are the top selling organic category, and many produce farmers use both direct and indirect channels. Furthermore, we focus on farmers who grow produce only and do not grow any field crops or livestock animals. We want to control for the potential spillover effects of marketing field crops or livestock to the marketing of fresh produce.

We start with a brief description of buyers' preferences in direct and indirect markets. Then we develop a framework that integrates production quality and marketing transaction costs. The framework provides a partial explanation of a farmer's choice of channels in marketing their output, and motivates the econometric specification, which is discussed in a later section. Based upon the analysis in the conceptual framework, we develop several hypothesis.

3.1 Background

We assume that transaction costs of marketing organic produce are channel specific and farm specific. Reflecting frictions in exchanges in the economic environment, transaction costs vary in kinds and magnitudes with characteristics of the market where the transaction occurs. Following Hobbs, we categorize transaction costs into information and search cost, negotiation costs, and monitoring and enforcement costs. Some transaction costs are fixed as they are invariant with quantities of exchange, and others are variable as they vary with quantities of exchange.

Direct markets are direct to consumers, including farmers market and community supported agriculture or subscription. Indirect markets are retails and wholesales, where retails include local supermarkets, natural food stores and restaurants and wholesales include distributors, processors and packers and handlers. Direct and indirect markets differ in a number of ways that would affect farmers' marketing choices.

Buyers in direct markets value freshness of produce and convenience of shopping. They prefer varieties and are heterogeneous in quality and other qualitative attributes such as size, color, shape and weight. Information on direct markets such as farmers' markets is readily available. Virtually any organic farmer can access farmer's markets with a small fee. Direct markets are usually concentrated in dense populations areas while farms are remotely located. Transportation costs to farmer markets are often substantial.

Buyers in indirect markets have distinctive preferences because of the nature of their business. The business model of retail and wholesale is high throughput rates; that is, they operate by moving a large quantity of homogeneous goods from producers to consumers quickly. Stated in another word, they prefer large quantity and consistency of quality. This business model is applicable where the retails or wholesalers carry conventional produce only, or organic produce only, or both.

3.2 Proportions versus Quantities

As stated earlier, this paper focuses on identifying the major transaction costs in marketing organic produce, and measuring their relative importance in influencing organic farmers' choice of marketing channels between direct and indirect markets. We are particularly interested in the effects of fixed transaction costs of entering indirect markets, because lowing the barriers to entry to retail and wholesale markets is critical to the organic market expansion. Consequently we want to explain organic farmers' use of indirect channels.

Two natural choices of the dependent variable are proportions – what proportion of total output is sold to indirect channels, and quantities – how many acres of total output are sold to indirect channels. The choice between proportions and quantities is an economic issue rather than a statistical issue. The question is: would farmers with *5* acres and *100* acres respond in the same way to changes in exogenous factors in terms of changes in the absolute size or relative share of total output sold to the indirect markets?

Consider two reduced form relationships: $Y = F(X, e)$ and $y = G(X, e)$, where Y is the number of acres sold to indirect markets, y is the proportion of the total production sold to indirect markets, X is the set of exogenous factors, and e is the random term. We are interested in the mean responses of changes in X. Consider a reduction of delivery distance from farm gates to wholesalers, would farmers with 5 acres and 100 acres both increase *1* acre of sales to wholesalers? or would their sales to wholesalers both increase by 10%? For a given type of produce at a given point of time, it is likely that retailers and wholesalers have much greater demand that what individual farmers can supply. Consequently, it is reasonable to model the mean response of the exogenous factors in proportions to indirect markets.

To reiterate the economic nature of the choice of the dependent variable, it is to be noted that proportions would be a poor choice if we were studying the farmer's use of direct markets, though it may be statistically valid to do so. It would be far-stretching to presume for the same change in the exogenous factors, a farmer with 5 acres would increase sales by 20 % or one acre and a farmer with 100 acres would increase sales by 20 % or 20 acres through roadside stands.

3.3 A Framework of Quality and Market Selection

The quality of fresh vegetable and fruits refers to the level of desirable qualitative characteristics of the produce, namely, nutritional quality, taste such as flavor and texture, and appearance such as size, shape, color and speckles. Because of the inherent variability of the biological processes in agricultural production, the quality of output exhibits a distribution density function. The presence of random environmental effects aggravates the quality variability.

The quality distribution of output depends on the farmer's skill and experience, quality of seeds and other inputs, soil and other natural resources, effort in organic pest and crop disease management and weed control, and random effects. It is thus appropriate to denote quality distributing as $f(q; w, \varepsilon)$ where q is the quality of output, w is a vector of farm- and farmer-specific exogenous factors and ε is the random shock. The support of $f(q; w, \varepsilon)$ is $[\underline{q}, \overline{q}]$.

We make some assumptions on the demand. We assume that retail and wholesale markets impose a quality requirement, such as a cut-off point of quality, denoted as q_0, such that only produce with quality q_0 or better can be marketed to retail and wholesale. In addition, we assume a uniform price in indirect markets. Stated in another word, indirect markets offer a single price p_0 for all quality above q_0, and no price premiums for higher quality than q_0.

We assume there is sufficient heterogeneity in buyers in direct markets such that quality of all levels are marketable to consumers directly. To simplify the analysis, we assume a two-tier pricing structure in direct markets: consumers pay $\underline{p}$ for all quality levels between $\underline{q}$ and q_0, and $\overline{p}$ for all quality levels between q_0 and $\overline{q}$. Further we assume $\underline{p} < p_0 < \overline{p}$. All farmers take prices as given, and receive the same prices regardless of quantities sold.

Let y^{id} and y^d denote the proportion of output sold to indirect and direct markets, respectively. In the absence of marketing transaction costs, a farmer would sell all his produce to the direct markets, $y^{id} = 0$ and $y^d = 1$, as a direct implication of our assumption $\underline{p} < p_0 < \overline{p}$.

Now consider transaction costs of selling to both markets in terms of fixed and variable transaction costs. Normalize the fixed transaction costs to direct market to zero. Let T denote the fixed transaction costs to indirect markets. Fixed transaction costs are farm-specific, and thus is denoted as $T(z, w)$ where w is a vector of farm- and farmer-specific characteristics as in the quality distribution $f(q; w, \varepsilon)$, and z is a set of transaction costs variables.

For simplicity, we assume constant variable costs. Let t^{id} and t^{d} denote the per-unit transaction costs in indirect and direct markets, respectively. Variable costs depend on the quality of produce marketed, as better quality produce may require better care and hence higher costs. Variable costs also depend on farm characteristics such as distance between the farm and the delivery points, whether the farmer owns or rents transportation vehicles, etc. Denote those characteristics as w[xiii], we write the variable costs as $t^{id}(q; z, w)$ and $t^{d}(q; z, w)$ where q is the quality level. Let τ denote the variable transaction cost differential at the point q where $\tau(q; z,w) = t^{d}(q; z,w) - t^{id}(q; z,w)$.

We assume farmers solve sequential optimization problem rather than multi-period dynamic optimization; we make this simplification because we do not have a panel dataset to take into account of the dynamics. Furthermore, we assume the per-period profit function is additively separable in marketing channels, that is, there are no spillover effects between channels during the current period.

We want to point out that the per-period separablity assumption does not preclude the effects of outcomes from previous periods. It is conceivable that a farmer who has been successful in consumer markets would have lower transaction costs entering the indirect markets. Scale and reputation are two possible sources of this kind of cross-channels spillover. The scale effect would lower variable transaction costs in both channels and the reputation effect would lower fixed transaction costs in indirect markets. We account for those lagged effects by denoting the current period's transaction costs as functions of exogenous factors including farm characteristics. This is consistent with the theoretical model because variables of previous periods are pre-determined in static optimization problems.

Under the above assumptions, a profit-maximizing price-taking farmer solves the following static optimization problem:

$$\pi^* = \max\left\{\pi^{id}, \pi^{d}, \pi^{b}\right\} \text{ where}$$

[3.11a] $$\pi^{id} = \int_{q_0}^{\bar{q}} \left[p_0 - t^{id}(s;z,w)\right] f(s;w,\varepsilon)ds + \int_{\underline{q}}^{q_0} \left[\underline{p} - t^{d}(s;z,w)\right] f(s;w,\varepsilon)ds - T(z,w)$$

[3.11b] $$\pi^{d} = \int_{q_0}^{\bar{q}} \left[\bar{p} - t^{d}(s;z,w)\right] f(s;w,\varepsilon)ds + \int_{\underline{q}}^{q_0} \left[\underline{p} - t^{d}(s;z,w)\right] f(s;w,\varepsilon)ds$$

[3.1c] $$\begin{aligned}\pi^{b} = \max_q &\int_{q_0}^{q} \left[p_0 - t^{id}(s;z,w)\right] f(s;w,\varepsilon)ds + \int_{q}^{\bar{q}} \left[\bar{p} - t^{d}(s;z,w)\right] f(s;w,\varepsilon)ds \\ &+ \int_{\underline{q}}^{q_0} \left[\underline{p} - t^{d}(s;z,w)\right] f(s;w,\varepsilon)ds - T(z,w)\end{aligned}$$

Optimization problem [3.1a] – [3.1c] reflects the fact that a farmer faces three discrete choices: (a) sell all high quality produce (produce that has quality q_0 or above) to indirect markets, and the profit of this choice is denoted as π^{id}, (b) sell all produce to direct markets, and the associated profit is denoted as π^{d}, and (c) sell to both channels by choosing the optimal allocation between the two channels, and the resulting profit is denoted as π^{b}.

Profit function π^{id} contains three terms: profit of selling all high quality produce to indirect markets, profit of selling the remaining low quality produce (produce that has quality below q_0) to direct markets, and the fixed transaction cost of entering indirect markets.

Profit function π^d contains two terms: profit of selling all high quality produce to the high-end direct markets where consumers pay $\overline{p}$, and profit of selling the remaining low quality produce to the low-end direct markets where consumers pay $\underline{p}$.

Profit function π^b involves an additional choice variable. Let q^* be the solution of [3.1c]. The maximand of [3.1c] contains four terms: profit of selling to indirect markets, profit of selling the best quality produce to high end direct markets, profit of selling the low quality produce to low end direct markets, and the fixed transaction costs of entering indirect markets.

Assuming an interior solution to maximization problem [3.1c], we characterize the solution by the following first-order and second order condition:
[3.2]

$$\left.\frac{\partial \pi^b}{\partial q}\right|_{q^*} = \left[p_0 - t^{id}(q^*;z,w)\right] f(q^*;w,\varepsilon) - \left[\overline{p} - t^d(q^*;z,w)\right] f(q^*;w,\varepsilon) = 0$$

$$\left.\frac{\partial^2 \pi^b}{\partial q^2}\right|_{q^*} = \left[t_q^d(q^*;z,w) - t_q^{id}(q^*;z,w)\right] f(q^*;w,\varepsilon) + \left[t^d(q^*;z,w) - t^{id}(q^*;z,w)\right] f_q(q^*;w,\varepsilon) < 0$$

If the quality density at q^* is non-zero, the first order condition can be written as:

[3.3] $$t^d(q^*;z,w) - t^{id}(q^*;z,w) = \overline{p} - p_0$$

which means the marginal profit of selling to either channel is the same. Solving [3.3] for q^*, we have an expression of q^* as a function of z and w, and consequently, the proportion sold to indirect markets, y^{id}, is a function of z and w:

[3.4] $$q^* = q^*(z,w)$$
$$y^{id} = \int_{q_0}^{q^*} f(q;w,\varepsilon)dq = \int_{q_0}^{q^*(z,w)} f(q;w,\varepsilon)dq$$

The effects of transaction costs on the optimal market selection can be described as follows:

[3.5]
$$\pi^* = \pi^d \Leftrightarrow \pi^d > \max(\pi^b, \pi^{id}) \Leftrightarrow T \geq \max\left\{\delta(q_0,q^*), \delta(q_0,\overline{q})\right\}$$
$$\pi^* = \pi^b \Leftrightarrow \pi^b > \max(\pi^d, \pi^{id}) \Leftrightarrow T < \delta(q_0,q^*) \;\&\; \delta(q^*,\overline{q}) < 0$$
$$\pi^* = \pi^{id} \Leftrightarrow \pi^{id} > \max(\pi^b, \pi^d) \Leftrightarrow T < \delta(q_0,q^*) \;\&\; \delta(q^*,\overline{q}) > 0$$

where $\delta(q_1,q_2) = \int_{q_1}^{q_2} [p_0 - t^{id}(q;z,w)] f(q;w,\varepsilon)dq - \int_{q_1}^{q_2} [\overline{p} - t^d(q;z,w)] f(q;w,\varepsilon)dq$. Condition [3.5] formalizes two main results. First, fixed and variable transaction costs jointly determine the entrance into indirect markets. When T is sufficiently high, or when the profit differential between indirect markets and high end direct market is sufficiently low, a farmer is rationed out of indirect markets and sells all output to direct markets.

Second, the proportion sold to indirect markets, conditional upon obtaining the access to indirect markets, is determined by variable transaction costs, and other exogenous factors such as prices and quality distribution. Specifically, fixed transaction costs do not affect the proportion sold to indirect markets, conditional on the market access [xiv].

We summarize the conditions for three observed choices – $y^{id} = 0$, $0 < y^{id} < 1$, and $y^{id} = 1$ as follows:

[3.6a] $$y^{id} = 0 \Leftrightarrow \pi^* = \pi^d \Leftrightarrow T \geq \max\left\{\delta(q_0,q^*), \delta(q_0,\overline{q})\right\}$$

[3.6b] $$y^{id} = 1 \Leftrightarrow \pi^* = \pi^{id} \;\&\; F(q_0) = 0 \Leftrightarrow T < \delta(q_0,q^*) \;\&\; \delta(q^*,\overline{q}) > 0 \;\&\; F(q_0) = 0$$

[3.6c] $$\begin{aligned} & y^{id} \in (0,1) \Leftrightarrow \pi^* = \pi^b \cup \langle \pi^* = \pi^{id} \ \& \ F(q_0) \neq 0 \rangle \\ & \Leftrightarrow \langle T < \delta(q_0, q^*) \ \& \ \delta(q^*, \bar{q}) < 0 \rangle \cup \langle T < \delta(q_0, q^*) \ \& \ \delta(q^*, \bar{q}) > 0 \ \& \ F(q_0) \neq 0 \rangle \end{aligned}$$

where F is the cumulated density function. Conditions [3.6b] and [3.6c] also make it explicit the role of quality distribution in determining the extreme cases $y^{id} = 0$ and $y^{id} = 1$. A very low quality production whereby $\int_{q_0}^{\bar{q}} f(q; z, \varepsilon) dq = 0$ is sufficient to determine $y^{id} = 0$. A very high quality production whereby $\int_{\underline{q}}^{q_0} f(q; z, \varepsilon) dq = 0$ is necessary to lead to the other extreme case $y^{id} = 1$.

3.4 Comparative Static

In the comparative static below, we denote z as the set of transaction cost variables, and z_1, z_2 as individual transaction cost variables; we denote w as the set of farm and farmer characteristics variables, and w_1, w_2 as individual characteristics variables.

3.4.1 Probability of Market Access

First, we consider the comparative static for market access, that is, how do z's and w's affect the probability of making a positive sales to indirect channels. Based upon the decision rules described in [3.6], we have:

[3.7a] $$\frac{\partial \Pr(y^{id} > 0)}{\partial z_1} = \underbrace{\frac{\partial \Pr(y^{id} > 0)}{\partial T(z,w)}}_{-} \frac{\partial T(z,w)}{\partial z_1} \propto -\frac{\partial T(z,w)}{\partial z_1}$$

[3.7b] $$\frac{\partial \Pr(y^{id} > 0)}{\partial w_1} = \underbrace{\frac{\partial \Pr(y^{id} > 0)}{\partial T(z,w)}}_{-} \frac{\partial T(z,w)}{\partial w_1} \propto -\frac{\partial T(z,w)}{\partial w_1}$$

[3.7c] $$\begin{aligned} \frac{\partial^2 \Pr(y^{id} > 0)}{\partial z_1 \partial w_1} &= \frac{\partial}{\partial w_1}\left(\underbrace{\frac{\partial \Pr(y^{id} > 0)}{\partial T(z,w)}}_{-} \frac{\partial T(z,w)}{\partial z_1} \right) \\ &= \underbrace{\frac{\partial}{\partial w_1}\left(\frac{\partial \Pr(y^{id} > 0)}{\partial T(z,w)} \right)}_{(1)=0} \underbrace{\frac{\partial T(z,w)}{\partial z_1}}_{(2)} + \underbrace{\frac{\partial \Pr(y^{id} > 0)}{\partial T(z,w)}}_{(3)<0} \underbrace{\frac{\partial}{\partial w_1}\left(\frac{\partial T(z,w)}{\partial z_1} \right)}_{(4)} \\ &\propto -\frac{\partial}{\partial w_1}\left(\frac{\partial T(z,w)}{\partial z_1} \right) \end{aligned}$$

where the symbol $\propto$ stands for proportional to, in another word, $a \propto b$ means a has the same sign as b.

Three things need to be noted. First, the common term $\frac{\partial \Pr(y^{id} > 0)}{\partial T(z,w)}$ is negative, and this is a direct consequence of the first condition in [3.5], where the higher fixed transaction costs of entering indirect markets, less likely to enter indirect markets or sell any to indirect markets at all. Second, term (1) in [3.7c] is zero, that is, the marginal effect of fixed transaction costs on the probability of market access is independent of transaction costs or farm characteristics variables where we have $\frac{\partial}{\partial z_1}\left(\frac{\partial \Pr(y^{id}>0)}{\partial T(z,w)}\right)=0$. Third, for any *z* or *w* or a combination of the two, the marginal effect on the probability of market access is negatively related to the marginal effect on the fixed transaction costs.

3.4.2 Probability of Market Penetration

The next set of comparative static results concerns the marginal effect on the probability of (indirect) market penetration conditional upon market access. First, let we want to evaluate the comparative static for sign of $\delta(q^*,\overline{q},z,w)$ as follows, using w_1 as an example:

[3.8]

$$\begin{aligned}
\frac{\partial \delta(q^*,\overline{q},z,w)}{\partial w_1} &= \frac{\partial}{\partial w_1}\int_{q^*(z,w)}^{\overline{q}} p_0 - t^{id}(q;z,w) - \overline{p} + t^d(q;z,w)dq \\
&= -q^*_{w_1}(z,w)\underbrace{\left[p_0 - t^{id}(q^*;z,w) - \overline{p} + t^d(q^*;z,w)\right]}_{(1)=0\text{ by FOC}} \\
&\quad + \int_{q^*(z,w)}^{\overline{q}} \underbrace{t^d_{w_1}(q;z,w)}_{(2)} - \underbrace{t^{id}_{w_1}(q;z,w)}_{(3)}dq \\
&= \int_{q^*(z,w)}^{\overline{q}} \underbrace{t^d_{w_1}(q;z,w)}_{(2)} - \underbrace{t^{id}_{w_1}(q;z,w)}_{(3)}dq
\end{aligned}$$

Based upon the decision rules described in [3.6] and results in [3.8], we have:

[3.8a]

$$\begin{aligned}
\left.\frac{\partial \Pr(y^{id}=1)}{\partial z_1}\right|_{y^{id}>0} &= \underbrace{\frac{\partial \Pr(y^{id}=1)}{\partial \delta(q^*,\overline{q},z,w)}}_{+}\frac{\partial \delta(q^*,\overline{q},z,w)}{\partial z_1} \\
&\propto \frac{\partial \delta(q^*,\overline{q},z,w)}{\partial z_1} \propto \int_{q^*(z,w)}^{\overline{q}} \underbrace{t^d_{z_1}(q;z,w)}_{(2)} - \underbrace{t^{id}_{z_1}(q;z,w)}_{(3)}dq
\end{aligned}$$

[3.8b]

$$\begin{aligned}
\left.\frac{\partial \Pr(y^{id}=1)}{\partial w_1}\right|_{y^{id}>0} &= \underbrace{\frac{\partial \Pr(y^{id}=1)}{\partial \delta(q^*,\overline{q},z,w)}}_{+}\frac{\partial \delta(q^*,\overline{q},z,w)}{\partial w_1} \\
&\propto \frac{\partial \delta(q^*,\overline{q},z,w)}{\partial w_1} \propto \int_{q^*(z,w)}^{\overline{q}} \underbrace{t^d_{w_1}(q;z,w)}_{(2)} - \underbrace{t^{id}_{w_1}(q;z,w)}_{(3)}dq
\end{aligned}$$

[3.8c]

$$\begin{aligned}
\left.\frac{\partial^2 \Pr(y^{id}=1)}{\partial z_1 \partial w_1}\right|_{y^{id}>0} &= \frac{\partial}{\partial w_1}\left(\underbrace{\frac{\partial \Pr(y^{id}=1)}{\partial \delta(q^*,\overline{q},z,w)}}_{+}\frac{\partial \delta(q^*,\overline{q},z,w)}{\partial z_1}\right) \\
&= \underbrace{\frac{\partial}{\partial w_1}\left(\frac{\partial \Pr(y^{id}=1)}{\partial \delta(q^*,\overline{q},z,w)}\right)}_{(1)=0}\underbrace{\frac{\partial \delta(q^*,\overline{q},z,w)}{\partial z_1}}_{(2)} + \underbrace{\frac{\partial \Pr(y^{id}=1)}{\partial \delta(q^*,\overline{q},z,w)}}_{(3)>0}\underbrace{\frac{\partial}{\partial w_1}\left(\frac{\partial \delta(q^*,\overline{q},z,w)}{\partial z_1}\right)}_{(4)} \\
&\propto \frac{\partial}{\partial w_1}\left(\frac{\partial \delta(q^*,\overline{q},z,w)}{\partial z_1}\right)
\end{aligned}$$

and

[3.8c′]
$$\frac{\partial}{\partial w_1}\left(\frac{\partial \delta(q^*,\bar{q},z,w)}{\partial z_1}\right)=\frac{\partial}{\partial w_1}\int_{q^*(z,w)}^{\bar{q}}\tau_{z_1}(q;z,w)dq$$
$$=-\frac{\partial q^*(z,w)}{\partial w_1}\tau_{z_1}(q^*;z,w)+\int_{q^*(z,w)}^{\bar{q}}\tau_{w_1 z_1}(q;z,w)dq$$

Three things need to be noted. First, the common term $\frac{\partial \Pr(y^{id}=1)}{\partial \delta(q^*,\bar{q},z,w)}=\frac{\partial \Pr(y^{id}=1)}{\partial \delta(z,w)}$ is positive, and this is a direct consequence of the second condition in [3.5], where the greater profit differential of selling high quality output to indirect versus to direct markets, the more likely the farm can sell all to indirect markets. Second, term (1) in [3.8c] is zero, that is, the marginal effect of profit differentials on the probability of market penetration is independent of transaction costs or farm characteristics variables where we have $\frac{\partial}{\partial z_1}\left(\frac{\partial \Pr(y^{id}=1)}{\partial \delta(q^*,\bar{q},z,w)}\right)=0$. Third, the sign [3.8c] and [3.8c′] depends on the sign of $\frac{\partial q^*(z,w)}{\partial w_1}$, which is discussed in the following subsection.

3.4.3 Amount of Indirect Sales

The following set of comparative static results [3.9a]-[3.9c] concerns the marginal effect on the magnitude of proportion conditional on the farmer is able to access but not penetrate the indirect markets. Based upon the decision rules described in [3.6], we have:

[3.9a]
$$\left.\frac{\partial y^{id}}{\partial z_1}\right|_{0<y^{id}<1}=\frac{\partial}{\partial z_1}\int_{q_0}^{q^*(z,w)}f(q;w,\varepsilon)dq$$
$$=\frac{\partial q^*(z,w)}{\partial z_1}\underbrace{f(q^*(z,w),w,\varepsilon)}_{+}+\int_{q_0}^{q^*(z,w)}\underbrace{f_{z_1}(q;w,\varepsilon)}_{0}dq\propto\frac{\partial q^*(z,w)}{\partial z_1}$$

[3.9b]
$$\left.\frac{\partial y^{id}}{\partial w_1}\right|_{0<y^{id}<1}=\frac{\partial}{\partial w_1}\int_{q_0}^{q^*(z,w)}f(q;w,\varepsilon)dq$$
$$=\frac{\partial q^*(z,w)}{\partial w_1}\underbrace{f(q^*(z,w),w,\varepsilon)}_{+}+\int_{q_0}^{q^*(z,w)}\underbrace{f_{w_1}(q;w,\varepsilon)}_{0}dq\propto\frac{\partial q^*(z,w)}{\partial w_1}$$

[3.9c]
$$\left.\frac{\partial^2 y^{id}}{\partial w_1\partial z_1}\right|_{0<y^{id}<1}=\frac{\partial}{\partial w_1}\left(\frac{\partial}{\partial z_1}\int_{q_0}^{q^*(z,w)}f(q;w,\varepsilon)dq\right)$$
$$\propto\frac{\partial}{\partial w_1}\left(\frac{\partial q^*(z,w)}{\partial z_1}f(q^*(z,w),w,\varepsilon)\right)$$
$$\propto\underbrace{\frac{\partial}{\partial w_1}\left(\frac{\partial q^*(z,w)}{\partial z_1}\right)}_{(1)}\underbrace{f(q^*(z,w),w,\varepsilon)}_{(2)>0}+\underbrace{\frac{\partial q^*(z,w)}{\partial z_1}}_{(3)}\underbrace{f_{w_1}(q^*(z,w),w,\varepsilon)}_{(4)}$$

Several things need to be noted. First, we use condition [3.4] to express the proportion y^{id} in terms of variables of interest, *z* and *w*. Second, the quality distribution of the output, *f(q; w, ε)* is independent of marketing transaction costs, that is, $f_z(q;\, w,\, \varepsilon)$ is zero. Third, conditions [3.9a]-[3.9c] state that all the marginal effects critically depend on the sign of the marginal effect on the optimal point q^*, $\frac{\partial q^*(z,w)}{\partial z_1}, \frac{\partial q^*(z,w)}{\partial w_1}$ or $\frac{\partial^2 q^*(z,w)}{\partial z_1 \partial w_1}$.

Before evaluating the signs of these terms, we need to make additional assumptions on the quality distribution at the optimal point q^*. We consider the cases where $f_q(q^*;\, w,\, \varepsilon)$ is positive, zero or negatively. Re-arranging the terms in the second order condition in [3.2], we obtain:

[3.10a]
$$\left[t_q^d(q^*;z,w) - t_q^{id}(q^*;z,w)\right] < -\frac{\overbrace{\left[t^d(q^*;z,w) - t^{id}(q^*;z,w)\right]}^{+} f_q(q^*;w,\varepsilon)}{\underbrace{f(q^*;w,\varepsilon)}_{+}}$$

$$t_q^d(q^*;z,w) - t_q^{id}(q^*;z,w) < 0,\ \text{if } f_q(q^*;w,\varepsilon) \ge 0,\ \text{undetermined o/w}$$

where in the first statement, the first term in the numerator at right hand side is positive because of condition [3.3] and assumption $\bar{p} > p_0$, and the denominator is positive because of the assumption we make in obtaining [3.3] from [3.2]. The second order condition tells us that the variable transaction costs differential between direct and indirect markets is negative if the slope of the quality distribution is non-negative at the optimal solution.

We apply the implicit function theorem on the first order condition in [3.2] and we have:

[3.10b]
$$\frac{\partial q^*(z,w)}{\partial z_1} = -\frac{\dfrac{\partial t^d(q^*,z,w)}{\partial z_1} - \dfrac{\partial t^{id}(q^*,z,w)}{\partial z_1}}{\dfrac{\partial t^d(q^*,z,w)}{\partial q^*} - \dfrac{\partial t^{id}(q^*,z,w)}{\partial q^*}}$$

Using [3.10a] and [3.10b] and assuming $f_q(q^*;\, w,\, \varepsilon) \ge 0$, we have:

[3.10c]
$$\frac{\partial q^*(z,w)}{\partial z_1} = -\frac{\dfrac{\partial t^d(q^*,z,w)}{\partial z_1} - \dfrac{\partial t^{id}(q^*,z,w)}{\partial z_1}}{\underbrace{\dfrac{\partial t^d(q^*,z,w)}{\partial q^*} - \dfrac{\partial t^{id}(q^*,z,w)}{\partial q^*}}_{-}} \propto \frac{\partial t^d(q^*,z,w)}{\partial z_1} - \frac{\partial t^{id}(q^*,z,w)}{\partial z_1}$$

and

[3.10d]
$$\frac{\partial q^*(z,w)}{\partial w_1} = -\frac{\dfrac{\partial t^d(q^*,z,w)}{\partial w_1} - \dfrac{\partial t^{id}(q^*,z,w)}{\partial w_1}}{\underbrace{\dfrac{\partial t^d(q^*,z,w)}{\partial q^*} - \dfrac{\partial t^{id}(q^*,z,w)}{\partial q^*}}_{-}} \propto \frac{\partial t^d(q^*,z,w)}{\partial w_1} - \frac{\partial t^{id}(q^*,z,w)}{\partial w_1}$$

Recall that $\tau(q;\, z,w) = t^d(q;\, z,w) - t^{id}(q,z,w)$. Substituting [3.10c] into [3.9a] - [3.9c], we have:

[3.9a′]
$$\left.\frac{\partial y^{id}}{\partial z_1}\right|_{0<y^{id}<1} \propto \frac{\partial \tau(q^*,z,w)}{\partial z_1}$$

[3.9b′]
$$\left.\frac{\partial y^{id}}{\partial w_1}\right|_{0<y^{id}<1} \propto \frac{\partial \tau(q^*,z,w)}{\partial w_1}$$

[3.9c′] $$\left.\frac{\partial^2 y^{id}}{\partial w_1 \partial z_1}\right|_{0<y^{id}<1} \propto \frac{\partial}{\partial w_1}\left(\frac{\partial q^*(z,w)}{\partial z_1}\right) \propto \frac{\partial^2 \tau(q^*,z,w)}{\partial w_1 \partial z_1}, \text{if } \left. f_q\right|_{q^*} \geq 0, \left. f_{w_1}\right|_{q^*} = 0$$

Condition [3.9c′] is obtained when we assume term (4) in [3.9c] is zero, that is, $f_{w1}(q^*(z,w),w,\varepsilon) = 0$, the density function is independent of w_1 at the optimal point $q^*(z,w)$, or a strong condition would require that the density function is independent of w_1 at any point of q.

In summary, condition [3.9a′]-[3.9c′] state that for any z or w or a combination of the two, the marginal effect on the magnitude of proportion conditional on the farmer is able to access but not penetrate the indirect markets is positively related to the marginal effect on the variable transaction cost differentials evaluated at the optimal solution.

Finally, we use [3.10a] and [3.10b] to determine the sign of [3.8c] and [3.8c′] as follows:

[3.8c″] $$\begin{aligned}\left.\frac{\partial^2 \Pr(y^{id}=1)}{\partial z_1 \partial w_1}\right|_{y^{id}>0} &\propto -\frac{\partial q^*(z,w)}{\partial w_1}\tau_{z_1}(q^*;z,w) + \int_{q^*(z,w)}^{\bar{q}} \tau_{w_1 z_1}(q;z,w)\,dq \\ &\propto -\tau_{w_1}(q^*;z,w)\tau_{z_1}(q^*;z,w) + \int_{q^*(z,w)}^{\bar{q}} \tau_{w_1 z_1}(q;z,w)\,dq\end{aligned}$$

3.4 Testable Hypothesis

We use the comparative static analysis to develop several hypotheses that we wish to investigate empirically. We wish to investigate how a transaction cost or a farm (or farmer) characteristics affects the probability of not being rationed out, $Pr(y^{id} > 0)$, the probability of penetration, $Pr(y^{id} = 1)$, and the proportion conditional upon not being rationed out and no penetration, $y^{id} \mid 0 < y^{id} < 1$.

Transaction costs considered include market infrastructure variables such as total number of distributors in the state, market condition variables such as lack of acceptance of organic certification and lack of consumer understanding, and specific transaction costs variables such as distance and reliable payment.

3.4.1 Hypothesis 1:Farming History Matters

We hypothesize that organic farmers who have no conventional farming history face substantially higher fixed transaction costs in entering the indirect markets than those who do, and as a result, they are less likely to sell to indirect channels. Analytically, let w_1 be TRANSORB (transitioner or beginner), an indicator variable whether the farmer has conventional farming experience ($w_1 = 1$) or not ($w_1 = 2$), substituting into [3.7b], we have:

[3.11a] $$\frac{\partial \Pr(y^{id}>0)}{\partial w_1} \propto -\underbrace{\frac{\partial T(z,w)}{\partial w_1}}_{-} > 0$$

This hypothesis is motivated by the nature of retailers and wholesalers business model. Logistics is the most important part of any retailing and wholesaling business, and more so for fresh produce because of the perishable nature of the produce. The profitability of retailers and wholesalers critically depends on supply channel relationships. Successful supply channel relationships mean low logistic and operational costs. As a consequence, retailers and wholesalers prefer stable long-term relationships with their suppliers.

Trust and reputation are essential to a long-term business partnership. Building trust and reputation, however, is a gradual and iterative process, as well as time- and resource-consuming. In addition, retailers and wholesalers do not prefer switching partners, and they would accept new partners only when the new suppliers have repetitively and demonstrably much better than the existing ones. This "stickiness" imposes additional barriers to entry to many organic farmers who try to make their inroads to the retail and wholesale markets.

We suppose organic farmers who transitioned from conventional farming could leverage the relationship they established through the conventional farming periods, that is, the farmer's personal trust and reputation can be carried over from conventional to organic production. Those who started organic farming from beginning and had no conventional farming experience would need to go through the highly competitive and costly relationship-building phase. In addition, new entrants are risk averse and they may trade quantity for quality, where they only sell the best quality of produce to demonstrate their reputation as a high quality supplier. This trade-off further reduces the proportions of their sales to the indirect markets.

In addition, we hypothesize that risk-averse retailers and wholesalers would prefer establishing long-term supply channel relationships with corporation farms, because corporations are more accountable than family-oriented partnership in cases of disputes, as the legal rules that confine a corporation's conduct provide a safeguard for its business partners. Analytically, let w_2 be BUSSTYPE, where a higher value indicates a more formal business structure, substituting it into [3.7b], we have:

[3.11b] $$\frac{\partial \Pr(y^{id} > 0)}{\partial w_2} \propto -\underbrace{\frac{\partial T(z,w)}{\partial w_2}}_{-} > 0$$

3.4.2 Hypothesis 2: Market Infrastructure and Market Condition

We hypothesize that better (indirect) market infrastructure and market environment would improve the probability of market access by lowering both the fixed transaction costs. Intuitively, measures of infrastructure and market condition can be thought of as positive supply shifters in farmers' access to indirect channels.

Analytically, let z_1 denote the degree of the lack of acceptance of organic certification documents in certain markets (LACKCERT), let z_2 be a measure of marketing infrastructure, such as number of distributors and retailers available (TAMND). Substituting z_1, *and* z_2 into [3.7a], we have:

[3.12] $$\frac{\partial \Pr(y^{id} > 0)}{\partial z_1} \propto -\underbrace{\frac{\partial T(z,w)}{\partial z_1}}_{+} < 0, \quad \frac{\partial \Pr(y^{id} > 0)}{\partial z_2} \propto -\underbrace{\frac{\partial T(z,w)}{\partial z_2}}_{-} > 0$$

Furthermore, we hypothesize that, conditional upon market access, distance to markets or delivery points and lack of consumers' understanding of organic food would adversely affect the probability of making all sales to the indirect markets. While market infrastructure can be thought of as positive supply shifters in market access, distance can be thought of as a negative supply shifter in market penetration.

Distance and market infrastructure are related, such that distance decreases with the measure of the marketing infrastructure. Sales to retail and wholesale channels is limited by the total number of organic retailers and wholesalers located at reasonable distances. Distance is an important factor in a farmer's profit function as fresh vegetables and fruits are highly perishable and easily damageable in some qualitative attributes such as appearance. To increase indirect sales, farmers have to reach more retail and wholesale markets located at further distance; as a result, farmers must incur higher costs of transportation, both because of longer travel distance and because of the additional costs, such as better packaging and cooling technologies, to ensure the quality of produce during the longer travel. Analytically, let z_3 denote the distance to markets (DISTANCE), substituting them into [3.8a], we have:

$$[3.13a] \qquad \left.\frac{\partial \Pr(y^{id}=1)}{\partial z_3}\right|_{y^{id}>0} \propto \int_{q^*(z,w)}^{\bar{q}} \underbrace{t_{z_3}^{d}(q;z,w)}_{(1)} - \underbrace{t_{z_3}^{id}(q;z,w)}_{(2)}\,dq < 0$$

Particularly, we hypothesize that while distance, for example, increases variable costs of direct sales, it increases the variable costs of indirect sales by a greater amount, since it would cost more to transport higher quality (or more easily perishable or damageable) produce to distributors who have strict standards on quality inspection than to farmers' markets.

Similarly, we can apply the same argument for [3.13] to variable z_2 = TAMND, where having a greater number of distributors and wholesalers would reduce the variable costs of indirect sales, and have little change on the variable costs of indirect sales, formally, we have:

$$[3.13b] \qquad \left.\frac{\partial \Pr(y^{id}=1)}{\partial z_2}\right|_{y^{id}>0} \propto \int_{q^*(z,w)}^{\bar{q}} \underbrace{t_{z_2}^{d}(q;z,w)}_{(1)=0} - \underbrace{t_{z_2}^{id}(q;z,w)}_{(2)<0}\,dq > 0$$

3.4.3 Hypothesis 3: When Farming History Meets Market Characteristics

We hypothesize that organic farmers who do not have conventional farming history are more likely and more severely to be constrained by the market conditions than those who do. We use a similar argument as in the first hypothesis. Organic farmers who converted from conventional farming have already established long-term relationship with existing indirect channels, and can leverage such a relationship in marketing organic outputs. As a consequence, they are not sensitive to new additions of marketing infrastructure. However, organic farmers who have not roots in conventional farming must start from scratch, and consequently, the marginal benefit of increasing infrastructure would be high.

As in the first hypothesis, let w_1 = TRANSORB, an indicator variable whether the farmer has conventional farming experience ($w_1 = 1$) or not ($w_1 = 2$), let z_1 and z_2, be the marketing infrastructure measure (TAMND) and market condition (LACKCERT) as in the second hypothesis. Substituting them into [3.7c], [3.8c] and [3.9c′], we obtain:

$$[3.14] \qquad \frac{\partial^2 \Pr(y^{id}>0)}{\partial w_1 \partial z_k} \propto -\underbrace{\frac{\partial}{\partial w_1}\left(\frac{\partial T(z,w)}{\partial z_k}\right)}_{+} < 0, k=1,2$$

Similarly, we use [3.9c′] to investigate the interaction effect of farming history (w_1 = TRANSORB) and z_3 (distance to markets) on the probability of market penetration:

$$[3.15] \qquad \left.\frac{\partial^2 \Pr(y^{id}=1)}{\partial z_3 \partial w_1}\right|_{y^{id}>0} \propto -\underbrace{\tau_{w_1}(q^*;z,w)}_{(1)<0}\underbrace{\tau_{z_3}(q^*;z,w)}_{(2)<0} + \int_{q^*(z,w)}^{\bar{q}} \underbrace{\tau_{w_1 z_3}(q;z,w)}_{(3)<0}\,dq < 0$$

where term (1) is negative because we hypothesize that farmers who transitioned from conventional farming have smaller cost differentials between direct and indirect markets, as they have been optimized their business to indirect sales with conventional farming. Term (2) is negative for the same argument in [3.13]. Term (3) is negative because we assume that distance has a larger negative effect on the farmers who have no conventional farming experience than on those who do.

Conditions [3.14] and [3.15] imply that the effects of transaction costs on market selection are heterogeneous across the TRANSORB sample types. An appropriate empirical specification should allow the heterogeneity of treatment.

3.4.4 Hypothesis 4: Production Characteristics

Our fourth hypothesis concerns the supply side. As described above, retailers and wholesalers operate under the business model of high turnover rates, and their operation is designed and optimized for handling large quantities of homogeneous type of produce with consistent quality. Therefore, we hypothesize that specialized farms in terms of degree of homogeneity of output are favored in penetrating the retail and wholesale markets.

This hypothesis is motivated by the economy of scale and by the lack of economy of scope in variable transaction costs of marketing. Transactions in indirect markets involve large quantities, and average transaction cost to indirect markets is likely to decline with quantity marketed. It is not clear whether large farms sell more *shares* of their output to indirect markets; we suspect that it can hold if the economy of scale is sufficiently large. On the contrary, we suppose the economy of scope is minimal. Different kinds of produce may require different kinds of packaging, sorting and labeling. Additional quality inspection is required for each variety of produce marketed.

Analytically, let w_3 denote the number of variety of produce. Substituting it to [3.8b], we have:

[3.16] $$\left.\frac{\partial \Pr(y^{id}=1)}{\partial w_3}\right|_{y^{id}>0} \propto \int_{q^*(z,w)}^{\bar{q}} \underbrace{t^d_{w_3}(q;z,w)}_{(1)=0} - \underbrace{t^{id}_{w_3}(q;z,w)}_{(2)>0}\, dq < 0$$

where we assume that the variety of produce does not change the variable transaction cost at the direct markets such that term (1) is zero, but the variety of produce increases the variable transaction cost at the indirect markets (at all relevant quality levels) such than term (2) is positive. [3.16] shows that having more varieties of produce decreases the probability of selling all to indirect markets.

In addition, we investigate how the variety of produce affects the amount of indirect sales conditional on market access. We substitute w_3 into[3.9b′] and obtain:

[3.17] $$\left.\frac{\partial y^{id}}{\partial w_3}\right|_{0<y^{id}<1} \propto \frac{\partial \tau(q^*,z,w)}{\partial w_3} = \underbrace{t^d_{w_3}(q^*;z,w)}_{=0} - \underbrace{t^{id}_{w_3}(q^*;z,w)}_{>0} < 0$$

This is a special case of [3.16] whereby [3.16] is true for all relevant support of quality q, while [3.17] is evaluated only at a single point q^*.

3.4.5 Hypothesis 5: Dynamic Effect

As we discuss earlier, though our theoretical model is a single period optimization, we do not exclude the effects of outcomes from previous periods, such as farmers can benefit from the spillover effect from being successful in direct channels in previous years, they do so establishing a reputation of being a high quality, low cost and reliable suppliers to the retailers and wholesalers. Fundamental to organic marketing is building a relationship, which is a dynamic process. Building upon the arguments made in the first hypothesis, we hypothesize that there is a time effect, whereby the more years the farm has tried to sell to indirect markets, the more it can sell to indirect sales for the current year. We use the number of years that the farm has been certified organic, denoted as w_4, as a proxy to the length the farm has gone through in the relationship-building process. The time effect impacts the market access by effectively lowering the fixed transaction costs. Analytically, we substitute w_4 into [3.7b] and obtain:

[3.18]
$$\frac{\partial \Pr(y^{id} > 0)}{\partial w_4} \propto \underbrace{-\frac{\partial T(z,w)}{\partial w_4}}_{-} > 0$$

In addition, we wish to investigate the heterogeneity of time effects between the transitioners and beginners subsamples. We hypothesize that the beginners would benefit more from the spillover effect than the transitioners, that is:

[3.19]
$$\frac{\partial^2 \Pr(y^{id} > 0)}{\partial w_1 \partial w_3} \propto \underbrace{-\frac{\partial}{\partial w_1}\left(\underbrace{\frac{\partial T(z,w)}{\partial w_3}}_{+}\right)}_{+} > 0$$

4. Econometric Analysis

We start with a reduced form specification, and followed by a semi-reduced form specification. For each specification, we specify the model, estimate the parameters, discuss the estimates, and address certain specification issues. We consolidate our estimation results and discuss our finding, particularly the five hypotheses developed earlier. We close this section by pointing out the limitations in our econometric analysis and suggest areas for further work.

4.1 Reduced Form Evidence

Conditions [6a]-[6c] suggest that the observed proportion sold to indirect markets, y^{id}, can be expressed as a function of transaction costs, farm- or farmer-specific characteristics, and random shock to the production.

4.1.1 Specification

An important characteristics of the dependent variable is that it is censored at both an upper and lower limit. Nearly fifty percent (*178* out of *390* samples) observations are limit observations, with either *0* or *100* percent of produce sold to indirect markets. An econometric model that is appropriate for this kind of data is two-limit Tobit limited dependent variable model with maximum likelihood estimation. The dependent variable in the Tobit model is continuous. We check this assumption by visually examining the histograms of the dependent variable, which are shown in Figure 1. The histograms suggest that there is sufficient variation in the dependent variable within the limits so that it can be modeled as a continuous variable.

A two-limit Tobit model is specified as follows (Maddala):

[4.1]
$$y^* = x\beta + \mu$$
$$y = \begin{cases} L_1 & \text{if } y^* \le L_1 \\ y^* & \text{if } L_1 < y^* < L_2 \\ L_2 & \text{if } y^* \ge L_2 \end{cases}$$

where y^* is the latent variable, x is a vector of exogenous variables, and μ is the disturbance term independently distributed with zero mean and constant variance σ^2. β is the vector of parameters of interest. In this case, the lower limit $L_1 = 0$ and upper limit $L_2 = 1$.

The likelihood function for this model is:

[4.2] $$L(\beta,\sigma \mid y,x,L_1,L_2)=\prod_{y=L_1} F\left(\tfrac{L_1-x\beta}{\sigma}\right)\prod_{y=y^*}\frac{1}{\sigma} f\left(\tfrac{y-x\beta}{\sigma}\right)\prod_{y=L_2}\left[1-F\left(\tfrac{L_2-x\beta}{\sigma}\right)\right]$$

where f is the standard normal density function, F is the standard normal cumulative function, and three terms represent the product of the probabilities of the lower limit, non-limit, and upper limit observations, respectively.

Of interest are the conditional expectation and unconditional expectation of *y*, which are defined as follows:

[4.3a] $$E(y \mid L_1 < y^* < L_2) = x\beta + E(\mu \mid L_1 - x\beta < \mu < L_2 - x\beta) = x\beta + \sigma\,\tfrac{f_1-f_2}{F_1-F_2}$$

and

[4.3b] $$\begin{aligned} E(y) &= P(y=L_1)L_1 + p(L_1<y^*<L_2)E(y \mid L_1<y^*<L_2) + P(y=L_2)L_2 \\ &= F_1L_1 + x\beta(F_2-F_1) + \sigma(f_1-f_2) + (1-F_2)L_2 \\ &= x\beta(F_2-F_1) + \sigma(f_1-f_2) + (1-F_2) \text{ when } L_1=0, L_2=1 \end{aligned}$$

where $F_k = F\left({}^{L_k-x\beta}\!/\!_{\sigma}\right), f_k = f\left({}^{L_k-x\beta}\!/\!_{\sigma}\right), k=1,2$.

The marginal effect of the *i*-th element of *x* on *y* is defined as (Hobbs):

[4.4] $$\begin{aligned} \frac{\partial E(y)}{\partial x_i} &= \frac{\partial\left[x\beta(F_2-F_1)+\sigma(f_1-f_2)+(1-F_2)\right]}{\partial x_i} \\ &= (F_2-F_1)\hat\beta + x\hat\beta\left[\frac{\hat\beta}{\hat\sigma}f_1 - \frac{\hat\beta}{\hat\sigma}f_2\right] + \hat\sigma\left[-\frac{x\hat\beta}{\hat\sigma}\frac{\hat\beta}{\hat\sigma}f_1 - \frac{1-x\hat\beta}{\hat\sigma}\frac{\hat\beta}{\hat\sigma}f_2\right] + \frac{\hat\beta}{\hat\sigma}f_2 \\ &= (F_2-F_1)\hat\beta = prob(0<y^*<1)\hat\beta \end{aligned}$$

The marginal effects are the parameter estimates scaled by $(F_2 - F_1)$, the probability of the latent variable falling within the limits. Since $0 < (F_2 - F_1) < 1$ for any *x*, the marginal effects have the same sign as the parameter estimates, and only the magnitudes are downward scaled.

4.1.2 Estimation Equation

Substituting *x* for vectors of transaction costs and farm characteristics variables in the Tobit model [4.1], the following is our estimation equation:

[4.6] $$PROPID = \alpha + \sum_k \beta\, IC + \sum_l \gamma\, MKT + \sum_m \delta\, NC + \sum_n \lambda MC + \sum_r \varphi SOC + \mu$$

where dependent variable *PROPID* is the proportion of a farm's produce sold to indirect markets. *IC, NC* and *MC* are vectors of variables representing information, negotiation, and monitoring transaction costs, respectively. *k* = FINDMKTS, OBTACCS, and NOTFINDP; *m* = DISTANCE; *n* = RELPMT and FAILCOMM. MKT is a vector of variables of market characteristics; *l* = LACKCONS, LACKCERT, OVERSUP, TAMID and TAMD. SOC is a vector of socioeconomic and farm characteristics; *r* = BUSSTYPE, LABOR, ORGLAND, LANDOWN, YRSCERT, VARIETY, VALUEADD, TRANORB, FULLORP, TOTYRS, ORGYRS, AGE, EDUC and GENDER. μ is the error term, assumed to be identically independently (normally) distributed with mean zero and variance σ^2. α, β, γ, δ, λ, and φ are parameters to estimate.

We discuss the exogeneity of variables selected in [4.6]. First, the ex ante transaction costs variables are usually exogenous, since these costs occur prior to actual transactions. We argue that variables FINDMKTS, OBTACCS, and NOTFINDP are exogenous to an individual farmer's allocation among channels. Prior to selling to various markets, farmers spend time and resources on market research, such as finding relevant markets and their characteristics (FINDMKTS). Prices being the most important characteristics of any market (NOTFINDP). After gathering and analyzing market information, farmers attempt to obtain access to the markets of interest by making contacts to the relevant personnel in the buying organizations (OBTACCS). Examples of such contacts are making inquiries and sending product samples for demonstration.

Second, while some negotiation costs can be endogenous, the variable chosen in [4.6] representing negotiation costs, DISTANCE, is assumed to be exogenous, as we assume that farmers take their farm location as well as the locations of the relevant markets as given for the given year's marketing decision. This is also consistent with our theoretical model where we assume the production is pre-determined.

Third, the ex post transaction costs can be endogenous and should be treated with great caution. We have two variables representing monitoring costs in [4.6], RELPMT and FAILCOMM. RELPMT measures the reliability and promptness of payments. Payments are usually instantaneous in direct markets as consumers pay on the point of purchase. Payments in indirect markets can be an issue where payments are usually delayed, and schedule of payments may be part of the contract if there is one. It is usually the characteristics of the buyers in indirect markets, such as the cash flow management practices, that result in particular patterns of payments to the farmers. In another word, individual farmers have no influence on how retailers and wholesalers would make their payments.

Variable FAILCOMM measures the overall impact of buyers failing to honor commitments. This variable may be endogenous, for example, it is plausible that it is simultaneously determined with the proportion of output sold to indirect markets. We will discuss and analyze the potential endogeneity of FAILCOMM in more details in a later section.

The market characteristics variables – LACKCONS, LACKCERT, OVERSUP, TAMID and TAMD, are exogenous as we assume no individual farmer has market power in any form in the theoretical model. This is also consistent with existing agricultural empirical studies on the lack of market power at the (conventional) farm gate.

The farm characteristics variables – BUSSTYPE, LABOR, ORGLAND, LANDOWN and YRSCERT, are exogenous to farmers' marketing decision, conditional upon their productions. LABOR is the number of managers and workers who are employed all year round, excluding the workers who are hired only seasonally, because the labor hired seasonally may be correlated with the marketing channel selection. The production related variables – VARIETY and VALUEADD, are assumed to be pre-determined to the market channel selection, a primary assumption in this paper.

Lastly, the farmer characteristics variables, such as TRANORB, FULLORP, TOTYRS, ORGYRS, AGE, EDUC and GENDER, are exogenous. A farmer's labor supply, as whether farming full-time or part-time, is assumed to be pre-determined to the marketing decision, though it may be correlated with the production decision. The exogeneity of FULLORP follows from our assumption of the pre-determinacy of production.

4.1.3 Estimation Results

The Tobit maximum likelihood estimators for [4.6] are presented in Columns 2 and 3 in Table 5 with coefficients estimates in column 2 and marginal effects in column 3. The results indicate that many transaction cost variables are statistically significant. Two of three variables of information costs are significant: finding organic markets (FINDMKTS) at *5 %* level and obtaining access to existing markets (OBTACCS) at *10 %* level. Representing negotiation costs, the transport cost which is measured as the distance between producer and market or delivery point is significant at *1 %* level. Both variables of monitoring costs are significant: failure of buyers to honor commitments (FAILCOMM) at *10 %* level and reliable or prompt payments (RELPMT) at *5 %* level. Furthermore, four out of five variables of market characteristics are found to be significant. Lack of consumers' understanding about organic food (LACKCONS), oversupply of legitimate organic products in existing markets (OVERSUP), and total available retailers and wholesales (TAMID) are significant at *5 %* level, and total available farmers' markets (TAMD) is significant at *10 %* level.

The results also indicate that several socioeconomic and farm characteristics variables are significant. Among the five variables for farm characteristics, two are significant: type of business structure (BUSSTYPE) at *5 %* level and years that the farm has been certified organic (YRSCERT) at *10 %* level, while labor, acreages of land farmed organically, and acreages of land owned are not significant. Both variables of production characteristics are significant: number of varieties of produce (VARIETY) at *1 %* level and number of value-added products (VALUEADD) at *5 %.* Lastly, two out of seven variables of farmer characteristics are significant: how the farmer started farming organically -- transitioning versus beginner (TRANORB) at *1 %* level and gender (GENDER) at *5 %* level.

Interpretation of the regression coefficients in a Tobit model differs from that in the ordinary least square regression because of the censoring nature of the dependent variable. The marginal effects of changes in individual explanatory variables are smaller than the coefficients, because they take into consideration of the probability of the sample's being within the lower and upper limits. The scaling factor of $Prob\ (0 < y^* < 1)$ is *0.4846*. The marginal effects are The magnitude of the scale factor depends on the number of observations that are at either limit. Our data has *49.3* percentage of observations that are censored, and this highly censored data effectively reduces the marginal effects to half of the regression coefficients.

4.2 Reduced Form Specification Issues

We discuss several specification issues in the Tobit model [4.6], namely, heteroskedasticity and clustering effect, independent omitted variables, simultaneity, and heterogeneity of effects.

4.2.1 Heteroskedasticity and Clustering Effect

The standard Tobit model [4.1] assumes homoskedasticity. Estimates of Tobit specification are sensitive to the assumption of homoskedasticity[xv]. We relax the assumption of homoskedasticity in two ways.

First, we allow for clustering effect where there is a pair-wise correlation such that $E(\mu_i\mu_j) \neq 0$. We still assumes independence across clusters. Clustering effect is likely since we include state-level data such as TAMND and TAMD as regressors. If there is positive (negative) clustering effect such that $E(\mu_i\mu_j) > 0$ ($E(\mu_i\mu_j) < 0$), standard errors of coefficient estimates would be biased downward (upward), and the magnitude of downward (upward) biases can be substantial.

In our estimation, we assume a random state effect. Let μ_{is} denote the error term of sample i in state *s*, we decompose μ_{is} into a state effect and an idiosyncratic component as follows:

[4.7] $$\mu_{is} = \alpha_s + \nu_{is} \text{ where } \nu_{is} \sim i.i.d.\ N(0,\sigma^2), \alpha_s \sim N(0,\sigma_s^2), E(x,\alpha_s) = 0$$

Second, we allow heteroskedasticity in addition to clustering effect. Heteroskedasticity is likely in many micro-level surveys. Our preliminary estimates in column (2) and (3) in Table 4 suggest that the variances of the transitioners and beginners subsamples differ systematically, when we assume homoskedasticity within the subsample. In account for heteroskedasticity, we use Huber and White robust variance estimators.

Accounting for both heteroskedasticity and clustering effect (by state of operation), we present the coefficients estimates and marginal effects in column (3) and (4) in Table 5[xvi].

4.2.2 Independent Omitted Variables

The second specification issue concerns with omitted variables that are independent of the explanatory variables x[xvii]. The issue of omitted variables arises when the choice of variables in our estimation is confined to the existing survey dataset. Suppose the true specification of the latent variable equation in [4.1] is the following:

[4.8] $$y^* = x\beta + \gamma c + \mu$$

where c is a scalar, and $\mu \mid x, c \sim Normal\ (0, \sigma^2)$, because the first element of x is the constant term, $E(c) = 0$ without loss of generality. Suppose that c is independent of x and has a distribution $c \sim Normal(0, \xi^2)$. Under these assumptions, the composite term $\gamma c + \mu \mid x \sim Normal\ (0, \gamma^2\xi^2 + \sigma^2)$. The likelihood function of model [4.8]is similar to [4.6]:

[4.9] $$L(\beta,\tau \mid y,x,L_1,L_2) = \prod_{y=L_1} F\left(\tfrac{L_1 - x\beta}{\tau}\right) \prod_{y=y^*} \frac{1}{\tau} f\left(\tfrac{y - x\beta}{\tau}\right) \prod_{y=L_2} \left[1 - F\left(\tfrac{L_2 - x\beta}{\tau}\right)\right]$$

where $\tau^2 = \gamma^2\xi^2 + \sigma^2$, and parameters of interest β remain unchanged from the original model [4.1]. Therefore, the omission of *independent* explanatory variables is harmless to our estimation.

4.2.3 Simultaneity

Unlike the independent omitted variables case, a serious specification issue arises when an explanatory variable is endogenous. The endogeneity can result from correlated omitted variables, measurement errors or simultaneity. Consider the case of simultaneity. Following Nelson and Olson, we suppose the latent variable equation in [4.1] is:

[4.10] $$\begin{aligned} y_1^* &= \alpha_1 y_2 + x_1\beta_1 + \mu_1 \\ y_2 &= \alpha_2 y_1^* + x_2\beta_2 + \mu_2 \end{aligned}$$

where y_2 is always observed, y_1* is observed only when it falls within the limits, and y_1 is censored at both limits and it is equivalent to y in [4.1], that is, the proportion of output sold to indirect channels. In this specification, y_2 is endogenous, and it depends on the latent variable y_1*, not the observed variable y_1. μ_1 and μ_2 are disturbance terms with a joint density $g(\cdot,\cdot)$ known up to a set of parameters θ. x_1 and x_2 are vectors of exogenous variables. α_1, α_2, β_1 and β_2 are the structural parameters to be estimated.

The likelihood function of specification [4.10] is as follows:

[4.11] $$\begin{aligned} &L(\alpha_1,\alpha_2,\beta_1,\beta_2,\theta \mid x_1,x_2,y_1,y_2) \\ &= \prod_{y_1=L_1} \int_{-\infty}^{L_1} g(y - \alpha_1 y_2 - x_1\beta_1, y_2 - \alpha_2 y - x_2\beta_2)\,dy \\ &\times \prod_{y_1=y_1^*} g(y_1 - \alpha_1 y_2 - x_1\beta_1, y_2 - \alpha_2 y_1 - x_2\beta_2) \\ &\times \prod_{y_1=L_2} \int_{L_2}^{\infty} g(y - \alpha_1 y_2 - x_1\beta_1, y_2 - \alpha_2 y - x_2\beta_2)\,dy \end{aligned}$$

The three terms present the product of the probabilities of the lower limit, within-limits and upper limit observations, respectively, similar to those in [4.2]. Maximizing [4.11] is computationally feasible when only are a small number of endogenous regressors [xviii].

The conditional and unconditional expectation of $y = y_1$ in model [4.10] are:

$$\begin{aligned} & E(y \mid L_1 < y^* < L_2) \\ \text{[4.12a]} \quad & = \alpha_1 y_2 + x_1\beta_1 + E(\mu_1 \mid L_1 - \alpha_1 y_2 - x_1\beta_1 < \mu_1 < L_2 - \alpha_1 y_2 - x_1\beta_1) \\ & = \alpha_1 y_2 + x_1\beta + \sigma_1 \tfrac{f_1 - f_2}{F_1 - F_2} \end{aligned}$$

and

$$\begin{aligned} & E(y) = P(y = L_1)L_1 + p(L_1 < y^* < L_2)E(y \mid L_1 < y^* < L_2) + P(y = L_2)L_2 \\ \text{[4.12b]} \quad & = F_1 L_1 + [\alpha_1 y_2 + x_1\beta][F_1 - F_2] + \sigma_1[f_1 - f_2] + (1 - F_2)L_2 \\ & = [\alpha_1 y_2 + x_1\beta][F_1 - F_2] + \sigma_1[f_1 - f_2] + (1 - F_2) \text{ when } L_1 = 0, L_2 = 1 \end{aligned}$$

where $F_k = F\left(\frac{L_k - \beta_1' x_1 - \alpha_1 y_2}{\sigma_1}\right), f_k = f\left(\frac{L_k - \beta_1' x_1 - \alpha_1 y_2}{\sigma_1}\right), k = 1,2$. [4.12a] and [4.12b] are equivalent to [4.3a] and [4.3b] where y_2 is not endogenous. Consequently, the marginal effect of the i-th element of x_1 or y_2 on y have the same form as [4.4]. Specifically,

$$\text{[4.13]} \quad \begin{aligned} \frac{\partial E(y)}{\partial x_{1i}} &= (F_2 - F_1)\hat{\beta}_1 = prob(0 < y^* < 1)\hat{\beta}_1 \\ \frac{\partial E(y)}{\partial y_2} &= (F_2 - F_1)\hat{\alpha}_1 = prob(0 < y^* < 1)\hat{\alpha}_1 \end{aligned}$$

where the marginal effects are the parameter estimates scaled by the probability that the latent variable falling within the limits.

An important characteristics of specification [4.10] is that the endogenous variable y_2 depends on the latent preference y_1^* rather than the observed variable y_1, consequently, the magnitude of the latent y_1^* directly affects the value of y_2. The intuitive appeal of this specification, argues Nelson and Olson, is that all else being equal, one might expect a different outcome for y_2 if y_1^* is way off the limits than when y_1^* is only marginally away from the limits.

We apply specification [4.10] to variable FAILCOMM. We want to take into account of the potential simultaneity of FAILCOMM on the parameters estimates of interest. Since we are not interested in the structural parameter of FAILCOMM, solving [4.10] for the reduced form, we obtain:

$$\text{[4.14]} \quad \begin{aligned} y_1^* &= x\beta + \upsilon_1 \\ y_2 &= x\pi + \upsilon_2 \quad \text{where } x = [x_1, x_2] \end{aligned}$$

β is the parameter vector of interest. In accounting for the endogeneity of variable FAILCOMM, we apply the reduced form solution [4.14] to the original estimation equation [4.6], by effectively eliminating the variable FAILCOMM from [4.6].

The corresponding estimation equation follows:

$$\text{[4.15]} \quad PROPID = \alpha + \sum_k \beta IC + \sum_l \gamma MKT + \sum_m \delta NC + \sum_n \lambda MC + \sum_r \varphi SOC + \mu$$

where dependent variable *PROPID* is the proportion of a farm's produce sold to indirect markets. n = RELPMT. The rest of variables and indexes are defined as the same as those in [4.6].

The parameter estimates and marginal effects are presented in Columns (5) and (6) in Table 5. The scaling factor $Pr\ (0 < y^* < 1)$ in [4.15] is *.4836*. Comparing the estimation results of [4.6] and [4.15], we see little changes in the parameter estimates as well as the marginal effects. Though this is not a formal exogeneity test, we believe the endogenity of variable FAILCOMM does not impose a serious problem[xix].

4.2.3 Heterogeneity of Effects

Specification [4.6] or [4.15], when estimated using the whole sample, assumes that the effects of transaction costs and other variables of interests are the same for the transitioners and beginners subsamples, after controlling for whether the sample is a transitioners or beginner.

A more flexible specification should allow for heterogeneous effects and test for whether the effects are homogeneous, rather than assuming the homogeneity in the specification. There are two approaches one can model heterogeneous effects in this setting. One approach is to interact every variable of interest in [4.15] with the indicator whether the sample is a transitioner or beginner (TRANSORB) and regress over the whole sample. The heterogeneity of effect is reflected in the coefficients on the interaction terms.

A second approach, which we use here, is to regress the specification [4.15] (with the absence of the term TRANSORB) on two subsamples separately. Table 6 presents the coefficients estimates and calculated marginal effects of the two subsamples. We find the results in Table 6 quite revealing. First, for all the transaction costs variable that are significant in the regression with the whole sample, if a particular transaction cost is significant for one subsample, it is not for the other subsample, and vice versa.

Second, while two transaction costs variables – finding markets and obtaining access, are significant for the transitioners, the beginners incur many more types of entrance barriers: distance, reliable payment, lack of consumer understanding, lack of acceptance of certification, over-supply, and total available indirect markets and direct markets. Particularly, all the market characteristics (LACKCONS, LACKCERT, OVERSUP, TAMND and TAMD) impose barriers to entry to the beginners, and *only* to the beginners.

Third, among the farm characteristics, business type is the only statistically significant factor for the transitioners and has a large coefficient, where a more formal or larger farm present an advantage. For the beginners, we find the size of land owned and how long the farm has been certified are both positively and significantly correlated with the access to the indirect markets, though the coefficients are quite small. For both subsamples, size of organic farming land has absolutely no explanatory power.

Fourth, the two subsamples do not differ much in the production characteristics. Produce variety is significant at *1%* and negatively correlated for both subsamples, and the marginal effects for both subsamples are similar. Number of value-added products are positively correlated but not significant for two subsamples.

Lastly, the error term of the beginners are smaller than that of the transitioners. This suggests that the beginners are more homogeneous than the transitioners.

We perform a Hausman test for the null hypothesis that all coefficients for the two subsamples do not differ systematically. The test statistics is $\chi^2(21) = 178.49$, and p-value = *.0000*. We hence reject the null. This lends an evidence that our specification should allow the heterogeneity of effects.

4.2.4 Measurement Errors

Measurement errors are common in micro-level data survey. They can cause considerable biases in the estimates, and in some cases, they can lead to spurious regression. In general, measurement errors in the dependent variable is less of a problem than measurement errors in the regressors. In our survey data, we believe the dependent variable is not much subject to measurement error in the first place – the respondents directly entered the proportion data, not the quantity data. In addition, the respondents entered the proportion data for a larger and more disaggregated set of markets, from which we aggregate into two broad categories: indirect and direct markets.

However, the transaction costs variables, directly taken from the survey, are likely subject to measurement errors for two reasons – these transaction costs variables are subjective and are categorical variables. Instrument variable approach is commonly used, however, not feasible in this study because of lack of data. We try to gauge the sensitivity of our estimates by simulating a particular measurement error. In the categorical data such as measured in *1* to *5* scale, a common measurement error is mis-categorization, where the respondent misplaced the data into the wrong category.

We perform one simple experiment as follows. For each transaction cost variable of interest, we randomly choose, say, *20* percent of the sample to increase the value by *1* and another *20* percent of the sample to decrease the value by *1* (keep the value intact if it is at the boundary, either *1* or *5*). Formally, we have the following:

$$\begin{aligned} &y = \alpha + \beta_1 x_1^* + \sum\nolimits_{i>1} \beta_i x_i + \sum\nolimits_j \gamma_j z_j + u \\ &x_1 = x_1^* + e_1, \text{where } E(e_1 \mid x_1^*) = 0 \\ &y = \alpha + b_1 x_1^* + \sum\nolimits_{i>1} \beta_i x_i + \sum\nolimits_j \gamma_j z_j + (b_1 e_1 + u) \end{aligned} \qquad [4.16]$$

where x_1 is the observed categorical value, x_1* is the unobserved true categorical value, and e_1 is the measurement error uncorrelated with the true value. Rest of the regressors have no measurement errors. This is the classical errors in variables case in a categorical variable. We have unbiased estimate of b_1 since the error term, $b_1 e1 + u$, has a conditional mean of zero. For a illustrative purpose, we choose e_1 to have three mass points at *–1*, *0*, and *1* with probabilities of *.2*, *.6* and *.2*, respectively, in our simple experiment.

We can extend to multiple regressors *x*'s as follows:

$$\begin{aligned} &y = \alpha + \sum \beta_i x_i^* + \sum\nolimits_j \gamma_j z_j + u \\ &x_i = x_i^* + e_i, \text{ where } E(e_i \mid x_{i,}^* \forall i) = 0 \\ &y = \alpha + \sum b_i x_i^* + \sum\nolimits_j \gamma_j z_j + (\sum b_i e_i + u) \end{aligned} \qquad [4.17]$$

where measurement errors *e*'s are uncorrelated with each other. We still get the unbiased estimates.

We present the estimates with simulated uncorrelated measurement errors in Table 8 using specification [4.15] for three samples separately. We make two comments while comparing Table 8 to Table 6. First, we find variable RELPMT, market infrastructure measures (TAMND and TAMD), business structure (BUSITYPE), and production characteristics variables (VARIETY and VALUEADD) are very robust both in terms of level of statistical significance and the magnitudes of estimates including coefficients and marginal effects. Similarly, we find variable DISTANCE and market condition variable OVERSUP are relatively robust.

Second, we find two ex ante transaction costs variables (FINDMKTS, OBTACCS) and two market condition measures (LACKCONS and LACKCERT) are sensitive to the simulated measurement errors, and sensitivity varies across the whole sample and two subsamples. Take OBTACCS as an example. Estimates with measurement errors increases the significance and magnitudes of the coefficients for the whole sample and the beginners subsample, but reduce the magnitudes of the coefficient for the transitioners subsample without changing the level of significance. Another example is LACKCONS and LACKCERT, where estimates of both variables with measurement errors decrease the significance and magnitudes of the coefficients for the whole sample and the beginners subsample, but reduce the magnitudes of the coefficient for the transitioners subsample without changing the level of significance.

We can improve several aspects in our simple experiment[xx]. First, the measurement errors of multiple variables can be correlated. Mis-categorization among various transaction costs variables may not be independent, as assumed in our first experiment. Modeling measurement errors is difficult, and modeling the correlations of multiple measurement errors is no less difficult. We look to the correlation between the data (transaction costs variables) reported, and use the same correlation matrix for the measurement errors. For example, if FINDMTKS and OBTACCS have a correlation of *.3*, then if FINDMTKS is randomly selected to be assumed to have mis-categorized, OBTACCS has a *.3* chance to be selected to be assumed to have mis-categorized in the same way.

Second, the measurement errors can occur *prior to* categorization. For a categorical variable x, which is scaled between *1* and *5*, for example, we can assume there is a continuous variable $\tilde{x}$,defined on the real line, such as:

[4.18] $$x = \begin{cases} 1 \text{ if } \tilde{x} \le 1 \\ k \text{ if } k < \tilde{x} \le k+1, k = 2,3,4 \\ 5 \text{ if } \tilde{x} > 5 \end{cases}$$

Suppose that $\tilde{x}$ has a measurement error e, such that $\tilde{x} = \tilde{x}^* + e$ where $\tilde{x}^*$ is the true value. In addition, we allow those measurement errors e's for multiple categorical variable x's are correlated.

<u>4.3 Semi-Reduced Form Evidence</u>

In previous subsections, we have estimated a Tobit model to examine the reduced form relationship between the proportion of indirect channels sales and a set of exogenous factors. We search for a more flexible econometric specification.

4.3.1 Motivations

Motivations for looking into an alternative econometric model are twofold. First, the economic model suggests that a farmer's optimal channel allocation is a multi-stage decision process where a farmer decides whether to enter indirect markets, and then allocates output between channels conditional upon market entrance. A graphical representation of the decision rules through conditions [3.5] are shown in Figure 2.

Second, the data is consistent with the multi-stage decision process. The data contains three types of farmers: sell all to indirect markets, sell all to direct markets, and sell to both markets. The histogram in Figure 1 shows that there are considerable concentration at the limit points. There is no prior reason to believe that the truncation at both limits points are symmetric, that is, the same exogenous factors affect the truncation at upper limit and lower limit in the identical way. The concentration and asymmetry suggest that a discrete-continuous model may be more appropriate than a single continuous model like the Tobit model.

The nature of multi-stage decision process raises additional econometric modeling issues: which factors determine market entrance? which factors determine the allocation between both markets conditional upon entering indirect markets? are these two sets of factors the same? do farmers who sell all to indirect markets respond to exogenous factors in the same way as farmers who sell to both markets?

If the factors that affect market entrance and channel allocation differ, or the effects of a particular exogenous factor that determines both market entrance and channel allocation differ in signs and / or magnitudes, the single equation reduced-form Tobit model would be inappropriate, because the single equation Tobit model assumes that the same exogenous factors determine both market entrance and channel allocation, and that all exogenous factors have the same effects on both market entrance and channel allocation. In this section, we want to relax those assumptions in the Tobit model and allow the heterogeneous effects of the exogenous factors.

An appropriate econometric model that corresponds to the decision process in Figure 2 is nested logit model. However, we are unable to apply the nested logit because of lack of "choice-specific" exogenous variables. Constrained by the dataset, we use a sequential logit model as the second best choice.

4.3.2 Specifications

Suppose that a set of exogenous variable x_1 determines the market entrance in the following way:

$$I_1^* = x_1\delta_1 + e_1 \qquad [4.19]$$
$$I_1 = \begin{cases} 1 & \text{if } I_1^* > 0 \\ 0 & \text{otherwise} \end{cases}$$

where I_1* is the latent variable; I_1 is observed, $I_1 = 1$ for $y^{id} > 0$ and $I_1 = 0$ for $y^{id} = 0$. Assuming the error term e_1 has a standard logistic distribution, the likelihood function for [4.19] is:

$$L(\delta_1 \mid y^{id}, x_1) = \prod\nolimits_{I_1} \left[G(x_1\delta_1)\right]^{I_1} \left[1 - G(x_1\delta_1)\right]^{1-I_1} \qquad [4.20]$$
$$\text{where } I_1 = 1 \text{ if } y^{id} > 0,\ I_1 = 0 \text{ if } y^{id} = 0$$
$$\text{Prob}(I_1 = 1 \mid x_1) = G(x_1\delta_1) = \Lambda(x_1\delta_1) \equiv \frac{\exp(x_1\delta_1)}{1 + \exp(x_1\delta_1)}$$

Similarly, we suppose that a set of exogenous variable x_2 determines whether a farmer sells all to indirect channels given indirect market entrance as follows:

$$I_2^* = x_2\delta_2 + e_2 \qquad [4.21]$$
$$I_2 = \begin{cases} 1 & \text{if } I_2^* > 0 \\ 0 & \text{otherwise} \end{cases}$$

where I_2* is the latent variable; I_2 is observed, $I_2 = 1$ for $0 < y^{id} < 1$ and $I_2 = 0$ for $y^{id} = 1$. Assuming the error term e_2 has a standard logistic distribution, the likelihood function for [4.21] is:

$$L(\delta_2 \mid y^{id} > 0, x_2) = \prod\nolimits_{I_2} \left[G(x_2\delta_2)\right]^{I_2} \left[1 - G(x_2\delta_2)\right]^{1-I_2} \qquad [4.22]$$
$$\text{where } I_2 = 1 \text{ if } y^{id} \in (0,1), I_2 = 0 \text{ if } y^{id} = 1$$
$$\text{Prob}(I_2 = 1 \mid x_2, y^{id} > 0) = G(x_2\delta_2) = \Lambda(x_2\delta_2) \equiv \frac{\exp(x_2\delta_2)}{1 + \exp(x_2\delta_2)}$$

Lastly, we suppose that a set exogenous variable x_3 determines the optimal channel allocation given that a farmer sells to both markets, that is, y^{id} is strictly between 0 and 1. A common approach is to assume that the log-odds ratio transformation, $\log [y^{id}/(1-y^{id})]$, has the conditional expectation of the form $x_3\,\delta_3$, such that the transformed dependent variable $\log [y^{id}/(1-y^{id})]$, ranges over the entire real line and parameters can be estimated using ordinary least squares. However, it is difficult to interpret the parameters δ_3, and thus it is difficult to recover the conditional expectation $E(y|\,x_3)$ that is of interest (Papke and Wooldridge).

Another approach that avoid this drawback is the fractional logit regression (Wooldridge p. 661) which models the conditional expectation $E(y|\,x_3)$ as a logistic function:

[4.23] $$E(y \mid x_3) = \frac{\exp(x_3\delta_3)}{1+\exp(x_3\delta_3)}$$

Seen as an extension of the binary logit model, the fractional logit model can be estimated using the quasi-MLE. The quasi-likelihood function of [4.23] has the same form as that of the binary logit model:

[4.24] $$L(\delta_3 \mid 0 < y^{id} < 1, x_3) = \prod\nolimits_{y^{id}} \left[G(x_3\delta_3)\right]^{y^{id}} \left[1 - G(x_3\delta_3)\right]^{1-y^{id}}$$ $$\text{where } G(x_3\delta_3) = \frac{\exp(x_3\delta_3)}{1+\exp(x_3\delta_3)}$$

Consequently, the estimation and inference of model [4.23] is similar to the binary logit model. For the obvious drawback of the log-odds transformation, we choose to use the fractional logit model [4.23] to explain the optimal channel allocation for those who use both channels.

4.3.3 Estimation

We estimate specifications [4.19], [4.21] and [4.23] independently and sequentially, and estimation results are presented in Table 9. First, the market entrance specification [4.19] is estimated using maximum likelihood estimator by maximizing the log of the likelihood function [4.20] with all observations available. The economic analysis through conditions [3.6] suggests that covariates of interest are fixed transaction costs variables, and factors that determine the profit differentials between direct and indirect channels for the high quality output. We include all transaction cost variables except the two direct market variables (lack of consumer understanding and total available direct markets), and we include all the socioeconomic characteristics.

We find that lack of acceptance of organic certification has the coefficient of *-.528*, largest magnitude among the transaction costs variables, and is significant at *1%,* which suggests it is the most severe Reliable payment and over-supply are positively correlated (*.377* and *.339*, respectively) and significant at *5%.* This is consistent with the tobit estimates in the earlier sections. The positive signs can be explained as follows: issues of payment and over-supply are likely to occur after the farm has an initial access to the indirect channels. The farmers who experience more problems with payment and competition are likely to be those who have gone further into the indirect channels.

Similarly, specification [4.21] is estimated using maximum likelihood estimator by maximizing the log of the likelihood function [4.22] with all non-zero observations, that is, $y^{id} > 0$. The choice of covariates is motivated from the decision rule in Figure *2*, which is derived from condition [3.6]. Covariates are chosen to represent the absence of low quality output and profit differentials between direct and indirect channels for the high quality output. We include all the transaction costs variables and socioeconomic characteristics. We find distance has a coefficient of *-.403*, largest magnitude among the transaction costs variables, and significant at *1%.* Reliable payment and lack of consumer understanding are also significant.

The optimal channel allocation specification [4.23], conditional upon both channels are used, is estimated using quasi-maximum likelihood estimator with all within limits observations, that is, $0 < y^{id} < 1$. Choice of covariates is motivated by condition [3.6]. We choose one transaction cost variable, namely, over-supply, and all the socioeconomic characteristics. We find that over-supply has a large coefficient of *.212* and is significant at *1%.*

Among the production variables, we find variety of produce is negatively correlated and is significant at *1%* in all three stages of estimation. Among the socioeconomic characteristics variables, being a beginner organic farmer has negative coefficients with large magnitudes and is significant in all three stages of estimation. This suggests that not coming from conventional farming is considerably disadvantaged.

As a validity check of this alternative specification, we include all variables as in the Tobit estimation in all three stages of logit estimation. We perform Hausman tests on the null hypothesis that the coefficients in stage l and coefficients in stage m are identical, where $l, m = 1, 2,3$ and $l \neq m$. We reject all the nulls[xxi].

4.3.4 Marginal Effects
The marginal effects in multi-stage logit estimation are obtained through simulation. We describe the mechanics of simulation as follows. Suppose we want to find the marginal effect of a transaction cost variable z, which is included in all three sets of covariates x_1, x_2 and x_3 through specifications [4.19], [4.21] and [4.23]. Furthermore, suppose we are interested in the change in y (proportion sold to indirect markets) on average by increasing one unit of z from its sample mean, holding all other covariates at their sample means. It makes sense to measure the change of mean y for one unit change of z since z is measured in integers (TAMID and TAMD variables) or *1* to *5* scale (all other transaction costs variables).

The simulation algorithm can be described as follows. For a given simulation size, for example, $N = 1,000$ and starting value of z at z_0, for example, let z_0 be the sample mean of z, in the first stage of simulation, we draw N logistically distributed random variables μ_1, calculate I_1* using specification [20] setting all other covariates at their sample means.

If $I_1* <= 0$, we set the $y = 0$, otherwise, we start the second stage of simulation as follows: we draw N logistically distributed random variables μ_2, calculate I_2* using specification [22] setting all other covariates at their sample means.

If $I_2* <= 0$, we set the $y = 1$, otherwise, we start the third stage of simulation as follows: we draw N logistically distributed random variables μ_3, calculate y_3 using specification [24] setting all other covariates at their sample means and set y to the average of y_3.. Then we take the average of y and denote it as y_0 as it is evaluated at z_0.

Now set z to z_1, for example, $z_1 = z_0 - 1$, repeat the above process, and denote the average of y as y_1. The difference of y_1 and y_0 is the marginal effect of z from z_0 to z_1 holding all other covariates at their sample means.

A selected set of simulated marginal effects of transaction costs variables are presented in columns (6) in Table 9. Compared with the marginal effects of the Tobit specification [4.15], which are presented in column (4) of Table 5, we make two observations: the marginal effects of all except one variable have the same signs; and the marginal effects in this alternative specification ([4.19], [4.21] and [4.23]) of all variables (except for the one which reverses the sign) have larger magnitudes than their counterparts in the Tobit specification [4.15].

We are interested in the sensitivity of our simulation results to the choice of simulation settings. Given the non-linear nature of logit models, we expect the simulation results to depend on a number of factors: choice of z, z_0, and z_1, choice of all other covariates denoted as x, and simulation size N. We present the marginal effects of a subset of transaction costs variables in Figure 3[xxii] based upon two dimensions of choices: the starting and ending value z_0, and z_1, and simulation size – *200*, *500* and *1,000*. The graph shows more room for irregularity than regularity.

4.4 Findings and Discussion

We summarize our findings from both reduced-form and semi-reduced form estimations, and we test the five hypothesis we have developed in section 3.4.

4.4.1 Do Transaction Costs Matter?

First of all, we want to show to what extend transaction costs explain market selection by testing the joint significance of all or a subset of transaction costs variables. In addition, we test whether transaction costs explain market selection equally well for the two subsamples. The primary goal of this study is to identify and measure the relative importance of marketing transaction costs on farmers' ability to sell to indirect channels.

We present all our hypothesis tests in Table 7, where the null hypothesis is all coefficients of variables indicated in the first column is zero. We make three comments. First, all transaction costs variables are jointly significant at 1% for all three samples. This suggests that transaction costs contribute to explaining farmers' indirect sales.

Second, how well certain subsets of transaction costs explain farmer's indirect sales vary dramatically and consistently between the transitioners and the beginner subsamples. From hypothesis (4)-(6) in Table7, we find market conditions (LACKCONS, LACKCERT and OVERSUP), market infrastructure (TAMND and TAMD), and the combination of both have no explanatory to the transitioners subsample, while they all do for the whole sample and the beginners subsample. So is the case for the remaining set of transaction costs (FINDMTKS, OBTACCS, NOTFINDP, DISTANCE and RELPMT). This lends the evidence that the transitioners and beginners face a very different set of transaction costs, and furthermore, the transaction costs the beginners face are more ubiquitous and more severe.

Third, test results for three sets of non-transaction costs variables – production characteristics (VARIETY and VALUEADD), farm characteristics (TOTYRS, ORGYRS, AGE, EDUC and GENDER) and farmer characteristics (BUSSTYPE, ORGLAND, LANDOWN, YRSCERT) are the same for all three samples at 1% significance level. We reject the null for the production characteristics. We fail to reject the nulls for the farm and farmer characteristics, although the p-values for the transitioners subsample is much greater than those for the whole sample or the beginners.

4.4.2 Hypothesis 1

Recall that hypothesis 1 says the probability of making any sales and the probability of selling all to indirect markets favor the transitioners and farmers with more formal business structure, that is, $\frac{\partial \Pr(y^{id}>0)}{\partial w_1} < 0, \frac{\partial \Pr(y^{id}>0)}{\partial w_2} > 0$ where w_1 = TRANSORB and w_2 = BUSSTYPE ([3.11a] and [3.11b]). We look for the evidence from the semi-reduced form estimates in Table 9. Coefficient of TRANSORB in [4.19] is -*.095* and significant at *1%.* Coefficient of BUSSTYPE in [4.19] is *.808* and significant at *5%.* Hence, we fail to reject hypothesis 1.

4.4.3 Hypothesis 2

Recall that hypothesis 2 has two components. We look for evidence from the semi-reduced form estimates. The first component says that lack of acceptance of organic certification presents an entrance barrier to the probability of making any sales to indirect markets, and better market infrastructure would improve the chance of market entrance, $\frac{\partial \Pr(y^{id}>0)}{\partial z_1} < 0, \frac{\partial \Pr(y^{id}>0)}{\partial z_2} > 0$ where z_1 = LACKCERT and z_2 =TAMND ([3.12]). Estimates in column (1) in Table 9 provide support for this where the coefficient of LACKCERT is -*.526* and significant at *1%,* and the coefficient of TAMND is *.011* and significant at *10%.*

The second component says distance to markets presents an obstacle in penetrating the indirect channel from the perspective of supply shifter. However, better market infrastructure would improve the chance of market penetration. That is, $\left.\frac{\partial \Pr(y^{id}=1)}{\partial z_3}\right|_{y^{id}>0} < 0, \left.\frac{\partial \Pr(y^{id}=1)}{\partial z_2}\right|_{y^{id}>0} > 0$ where z_3 = DISTANCE and z_2 = TAMND ([3.13a] and [3.13b]). Estimates of [4.21] in column (2) of Table 9 are consistent with this where the coefficient of LACKCONS is *.308* and significant at *5%,* the coefficient of DISTANCE is -*.403* and significant at *1%,* and the coefficient of TAMND is *.016* and significant at *10%.*

The reduced form estimates in column (1) of Table 6 provide further evidence. Coefficients of DISTANCE, LACKCONS and LACKCERT are .-*081*, -*.061* and -*.080*, respectively, and are significant at *1%*, *5%* and *1%,* respectively. Coefficient of TAMND is *.004* and significant at *1%.* Based upon both semi-reduced form and reduced form estimates, we fail to reject hypothesis 2.

4.4.4 Hypothesis 3

Recall that hypothesis 3 says that the beginners are more severely constrained by the market infrastructure (TAMND) and market condition (LACKCERT) in market entrance ([3.14]), and by distance to markets (DISTANCE) in market penetration, than the transitioners. We look for evidence from the semi-reduced estimates [4.15] in Table 6. Coefficient of TAMND is *.005* and significant at *1%* for the beginners, while it is *.002* and insignificant at *10%* for the transitioners. Coefficient of LACKCERT is -*.115* and significant at *1%* for the beginners, while it is insignificant at *10%* for the transitioners. Coefficient of DISTANCE is -*.073* and significant at *1%* for the beginners, while it is insignificant at *10%* for the transitioners. Based upon the reduced-form evidence, we fail to reject hypothesis 3.

4.4.5 Hypothesis 4

Recall that hypothesis 4 ([3.16] and [3.17]) says that having more varieties of produce reduces the probability of market penetration and reduces the amount of indirect sales conditional upon selling to both channels. That is, $\left.\frac{\partial \Pr(y^{id}=1)}{\partial w_2}\right|_{y^{id}>0} < 0, \left.\frac{\partial y^{id}}{\partial w_2}\right|_{0<y^{id}<1} < 0$, where w_2 = VARIETY.

Coefficient of VARIETY in logit estimates of [4.21] and [4.23] in Table 9 are -*.219* and -*.079*, respectively, and both are significant at *1%.* The reduced form estimate of VARIETY is -*.049* and significant at *1%.* Thus, we fail to reject hypothesis 4.

4.4.6 Hypothesis 5

Recall that hypothesis 5 says that there is a positive time effect in gaining market entrance, in another word, having been certified longer would improve the probability of making indirect sales ([3.18]). That is, $\frac{\partial \Pr(y^{id}>0)}{\partial w_4} > 0$ where w_4 = YRSCERT. Coefficient estimate of YRSCERT in [4.19] is *.115* and significant at *5%.* This suggests that there is a strong positive time effect, and success to indirect marketing is a gradual and dynamic process[xxiii].

In addition, hypothesis 5 says the beginners would benefit more from the dynamic process than the transitioners ([3.19]). We look for evidence in the reduced form estimates by the two subsamples in Table 6. The coefficients of YRSCERT for the transitioners and beginners are *.028* and *.022*, respectively, and the marginal effects of YRSCERT are *.011* and *.012*, respectively. Neither coefficient estimate is significant at *10%.* This does not provide ample evidence that time effect differs between the two subsamples.

4.5 Limitations and Extensions

We recognize a major limitation in our semi-reduced form estimation where we use the sequential logit model. A more appropriate model would be nested logit since decisions across stages can be correlated. For example, if there is an omitted variable, or an unobserved factor, such as risk preferences or reasons for adopting organic farming, may influence both market entrance decision and channel allocation decision. Without accounting for the correlation, out estimates may be biased. The direction and magnitude of biases, however, are difficult to determine without additional data.

In addition, as in the reduce form Tobit model, the transitioners and beginners subsamples may differ systematically. Hence, it would be desired to estimate the semi-reduced form model on both sub-samples separately. While technically feasible to do so, the sample size would be too small to make inferences from.

Our econometric analysis may be improved in the following ways. First, our treatment to measurement errors is far from satisfactory. We could experiment with more realistic cases to simulate the measurement error, and we can search for instruments to account for the measurement errors. Second, there may be over-sampling issues, for example, the beginners may be over-sampled. Also, our treatment to missing data (or item non-response) by discarding the observations may not be efficient.

We would like to extend our work in several ways. First, we could estimate the same models to a broader sample that include the organic produce farmers who also grow field crop and/or raise livestock animals. It would be interesting to compare the estimation results.

Second, we could use alternative econometric models and analyze a more disaggregated channel selection separating retail markets from wholesale markets. Estimators used in related literature include seemingly unrelated regression for multiple channels using the proportion data, and multinomial logit by converting the proportion data into dichotomy data. Other estimators may also be appropriate and provide new insights. One possibility is a system of equations where each equation is a two-limit Tobit regression.

Third, we could relax the assumption that production is pre-determined. Constrained by the lack of data on yields, costs and prices for individual crops for individual farmers, we won't be able to model production and marketing simultaneously or to analyze the supply responses. It is possible to model certain aspects of production jointly with the market selection. For example, the number of variety of produce is statistically significant in any of our specifications, we could possibly model the variety of produce and market selection simultaneously.

5. Conclusion

We develop a conceptual framework that integrates quality of output and transaction costs in the choice of marketing channels; based upon which, we estimate a reduced-form Tobit model and a semi-reduced logit model on a farm-level cross-sectional dataset.

We find strong empirical evidence that existing organic retail and wholesale markets impose considerable barriers to entry to individual organic farmers. Lack of marketing infrastructure such as small numbers of organic distributors, market condition such as lack of acceptance of organic certification, and prohibitively great distance between the farm gate to the markets resulted from immature market infrastructure are shown to be barrier to entry as well as penetrate to indirect channels.

We also find that the effects of transaction costs are asymmetric between the two types of farmers, those who transitioned from conventional farming and those who did not. Those who did are overall favored, and those who did not are constrained by more types of transaction costs and are constrained more severely than those who did. For example, lack of acceptance of organic certification is found to be the single most negative effect on those who did not, but statistically insignificant on those who did; similarly, market infrastructure adversely and significantly affects who those did not, but is insignificant on those who did.

Implications for potential organic agricultural policy are based upon our findings. While a policy that improves the overall market infrastructure to all organic farmers would encourage the market growth, an arguably more effective policy should target to specific segments of organic farmers, such as the least favored farmers those who did not transition from conventional farming. Discrimination where history of the organic farmers has the most decisive power in obtaining access to indirect markets, if found to exist, should be strongly discouraged. In another word, an effective policy should encourage or mandate distributors and retailers have a more transparent and objective process in selecting organic suppliers, such that all farmers would have an equal opportunity to be successful in selling to indirect markets.

References

Bailey, DeeVon and Lyn Hunnicutt, "The Role of Transaction Costs in Market Selection: Market Selection in Commercial Feeder Cattle Operations," paper presented at the Annual Meeting of the American Agricultural Economic Association in Long Beach, CA; July 28-31, 2002.

Dimitri, Carolyn and Catherine Greene. "Recent Growth Patterns in the U.S. Organic Food Market." Economic Research Service, U.S. Department of Agriculture, 2001, available at http://www.ers.usda.gov/publications/aib777/

Ferto, Imre and Gabor G. Szabo, "The Choice of Supply Channels in Hungarian Fruit and Vegetable Sector," paper for presentation to the American Agricultural Economics Association 2002, Annual Meeting, Long Beach, California.

Goetz, Stephan J., "A Selectivity Model of Household Food Marketing Behavior in Sub-Saharan Africa," *American Journal of Agricultural Economics* 74 (1992):444-452.

Greene, Catherine and Amy Kremen. "U.S. Organic Farming Emerges in 2000-2001: Adoption of Certified Systems." Economic Research Service, U.S. Department of Agriculture, 2003, available at http://www.ers.usda.gov/publications/aib780/

Hobbs, Jill E., "Measuring the Importance of Transaction Costs in Cattle Marketing," *American Journal of Agricultural Economics*, 79, (1997), 1083-1095

Key, Nigel, Elisabeth Sadoulet and Alain de Janvary, "Transaction Costs and Agricultural Household Supply Response," *American Journal of Agricultural Economics* 82 (2000):245-259

Klonsky, Karen and Laura Tourte, "Organic Agricultural Production in the United States: Debates and Directions," *American Journal of Agricultural Economics*, 80 (5) (1998) 1119-1124

Lampkin, N. H. and S. Padel, "Organic Farming and Agricultural Policy in Western Europe: An Overview," in *The Economics of Organic Farming*. N.H. Lampkin and S. Padel, eds. Wallingford UK: CAB International 1994.

Lampkin, Nicolas. "Development of Policies for Organic Agriculture," UK Organic Research 2002: Proceedings of the COR Conference, March 2002, Univ. of Wales, Aberystwyth.

Lee, Lung-Fei, G.S. Maddala, and R.P. Trost, "Asymptotic Covariance Matrices of Two-Stage Probit and Two-Stage Tobit Methods for Simultaneous Equations Models with Selectivity," *Econometrica* 48 (2) (1980):491-504.

Lohr, Luanne and Lennart Salomonsson, "Conversion Subsidies for Organic Production: Results from Sweden and Lessons for the United States," 1998, Working Paper, Department of Agriculture and Applied Economics, University of Georgia.

Lohr, Luanne, Yassert Gonzalez-Alvarez and Anita Graf, "Predicting Organic Market Development with Spatial Analysis of Existing Industry," working paper, University of Georgia, 2001.

Maddala G. S. *Limited Dependent and Qualitative Variables in Econometrics*. New York: Cambridge University Press, 1983.

McDonald, John F. and Robert A. Moffitt, “The Uses of Tobit Analysis,” *Review of Economics and Statistics*, 62 (1980), 318-321.

Nelson, F.D., and L. Olson, “Specification and Estimation of a Simultaneous Equation Model with Limited Dependent Variables.” *International Economic Review*, 19, (1978) 695-709.

Omano, Steve W., “Transport Costs and Smallholder Cropping Choices: An Application to Siaya District, Kenya,” *American Journal of Agricultural Economics*, 80, (1998) 116-123.

Pietola, Kyosti and Alfons O. Lansink, “Farmer response to policies promoting organic farming technologies in Finland,” *European Review of Agricultural Economics*, 28 (1) (2001) 1-15.

Rosett, Richard N. and Forrest D. Nelson, “Estimation of the Two-Limit Probit Regression Model,” Econometricia, 43 (1975) 141-146.

Walz, E., “Final Results of the Third Biennial National Organic Farmers’ Survey.” Organic Farming Research Foundation, Santa Cruz, California. Available online at http://www.ofrf.org/publications/survey/Final.Results.Third.NOF.Survey.pdf.

Williamson, Oliver E., *The Economic Institutions of Capitalism*. New York: Free Press. 1986.

Wooldridge, Jeffery M., *Econometric Analysis of Cross Section and Panel Data*. The MIT Press. 2002.

Figure 1: Histogram of observed choices

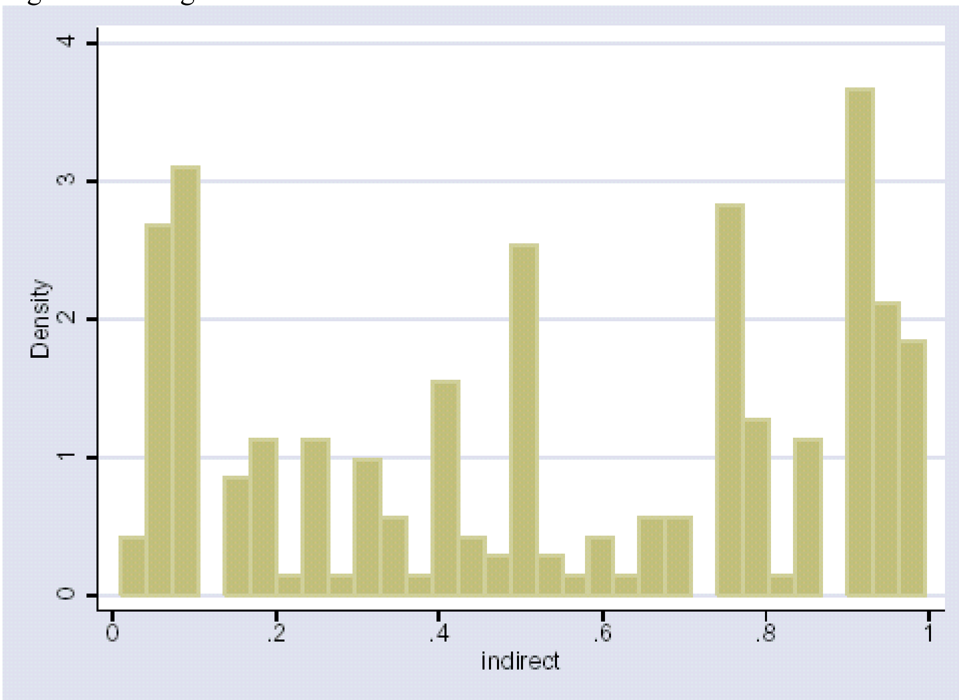

Figure 1a: Histogram of within limits observations where $y^{id} \in (0,1)$
N = 223, mean = 0.5272, standard deviation = 0.3335, min = 0.01, max = 0.995

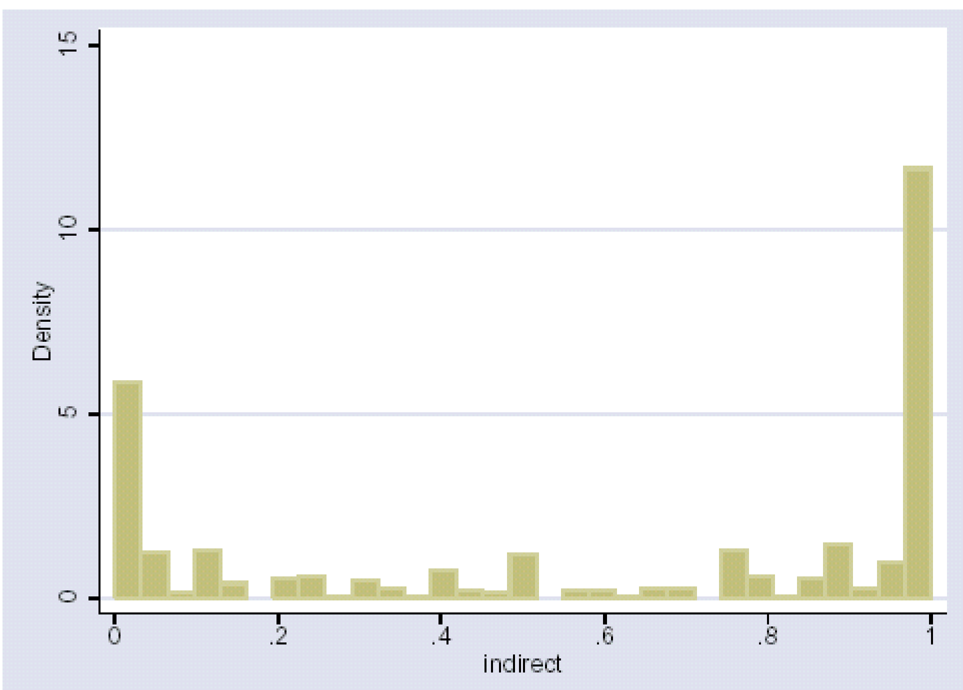

Figure 1b: Histogram of all observations including limit observations where $y^{id} \in [0,1]$
N = 471, mean = 0.5978, standard deviation = 0.4185, min = 0, max = 1

Figure 2: Decision tree representation of condition [3.6] for observed choices

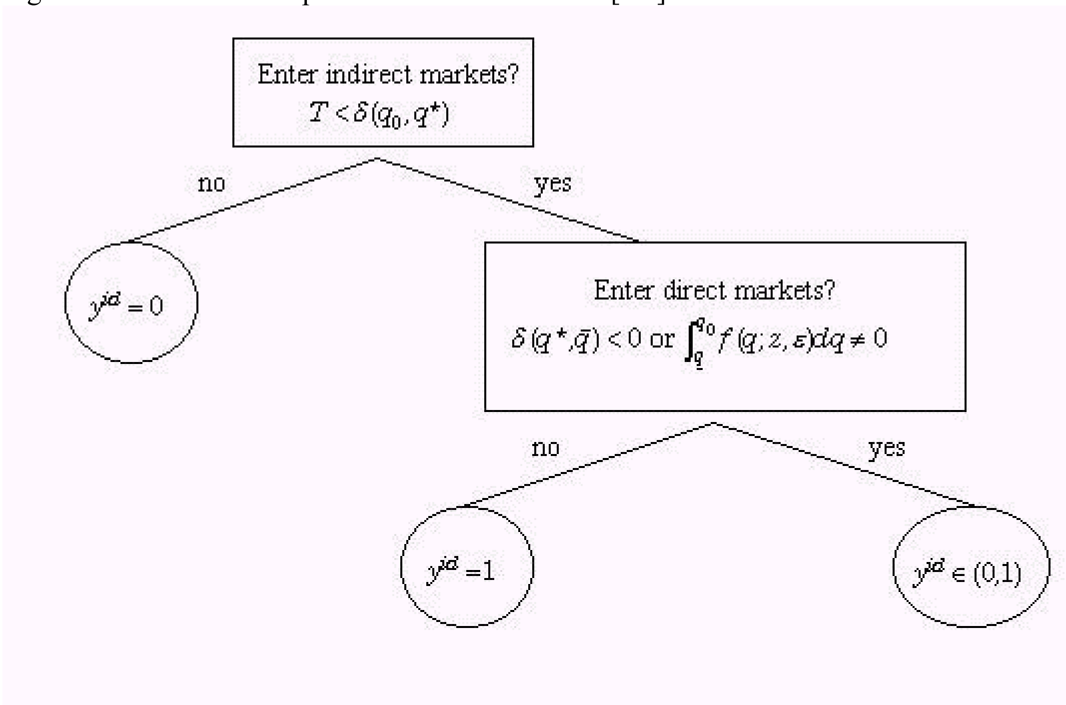

Table 1: Description of transaction cost variables

Variable Description	Variable Name	Measurement
Dependent Variable		
Proportion sold to indirect markets	PROPID	Proportion of produce sold into indirect markets (retail and wholesale)
Transaction Costs		
Information Costs		
Finding organic markets	FINDMKTS	How severe is "finding organic markets" a constraint to your marketing your organic products?
Obtaining access to existing markets	OBTACCS	How severe is "obtaining access to existing markets" a constraint to your marketing your organic products?
Inability to find best prices	NOTFINDP	How severe is "inability to find best prices" a constraint to your marketing your organic products?
Negotiation Costs		
Distance between producer and market or delivery points	DISTANCE	How severe is "distance between producer and market or delivery points" a constraint to your marketing your organic products?
Monitoring Costs		
Failure of buyers to honor commitment	FAILCOMM	How severe is "failure of buyers to honor commitment " a constraint to your marketing your organic products?
Reliable or prompt payment	RELPMT	How severe is "reliable or prompt payment" a constraint to your marketing your organic products?
Market Characteristics		
Lack of organic marketing network	LACKNWK	How severe is "lack of organic marketing network" a constraint to your marketing your organic products?
Lack of consumer understanding about organic food	LACKCONS	How severe is "lack of consumer understanding about organic food" a constraint to your marketing your organic products?
Lack of acceptance of certification documents in certain markets	LACKCERT	How severe is "lack of acceptance of certification documents in certain markets" a constraint to your marketing your organic products?
Oversupply of legitimate organic products in existing markets	OVERSUP	How severe is "oversupply of legitimate organic products in existing markets" a constraint to your marketing your organic products?
Total available indirect markets	TAMID	Number of retailers, processors and manufacturers, wholesalers in farm's state
Total available direct markets	TAMD	Number of farmer's markets in farm's state

Notes: Except for PROPID, TAMID and TAMD, all variables are measured in 1-5 scale at increasing severity level.

Table 2: Description of farm and farmer characteristics variables

Variable Description	Variable Name	Measurement
Farm Characteristics		
Type of business structure	BUSSTYP	Which of the following business structures describes your farm operation? [a]
All organic or mixed operation	ALLORMIX	Is all of your operation organic (1), or do you have a mixed organic and conventional operation (2)?
Hired labor	LABOR	Number of hired managers and workers [b]
Organic farming land	ORGLAND	How many acres do you farm organically?
Land owned	LANDOWN	How many acres do you own?
Years of the farm being certified organic	YRSCERT	How many years has your farm been certified organic?
Production Characteristics		
Number of varieties of vegetables and fruits	VARIETY	Number of types of vegetables and fruits produced and sold
Number of value-added products	VALUEADD	Number of types of value-added products produced and sold
Farmer Characteristics		
Transitioning or beginner?	TRANORB	Organic farmers can be classified as either starting from “scratch” as an organic producer (2) , or as “transitioning” from conventional agriculture (1). How did you start farming organically?
Full time or part time?	FULLORP	Do you farm full-time (1) or part-time (2)?
Total years of farming	TOTYRS	What is the total number of years you have been farming?
Years of farming organically	ORGYRS	How many years have you been farming organically?
Age	AGE	What is your age?
Education	EDUC	What is your level of formal education? [c]
Gender	GENDER	Your gender 1 = female, 2 = male

Notes: [a]: 1 = single family or family partnership, 2 = partnership other than family and cooperative, and 3 = corporation
[b]: We add up number of managers and workers hired, halved if part-time, halved if seasonal only.
[c]: 1 = no formal education, 2 = some high school, 3 = completed high school, 4 = some college, 5 = completed college, 6 = graduate work, and 7 = graduate degree.

Table 3: Summary statistics

Variable	All Sample Mean (SD) N = 390 (1)	Transitioners Mean N_1 = 111 (2)	Beginners Mean N_2 = 249 (3)	[t-stat] (4)
PROPID	.60 (.41)	.80	.51	2.36
FINDMKTS	2.34 (1.37)	2.58	2.24	.75
OBTACCS	2.33 (1.35)	2.59	2.27	2.64
NOTFINDP	2.78 (1.29)	2.91	2.72	1.63
DISTANCE	2.55 (1.42)	2.40	2.62	-1.07
FAILCOMM	1.96 (1.22)	1.96	1.96	1.55
RELPMT	2.09 (1.27)	2.13	2.06	2.35
LACKCONS	3.02 (1.38)	2.93	3.05	-4.46
LACKCERT	1.45 (.93)	1.51	1.42	1.02
OVERSUP	2.08 (1.27)	2.30	1.98	1.00
TAMID	22.91 (29.67)	25.45	21.78	0.05
TAMD	150.68 (126.12)	155.18	148.68	0.02
BUSSTYP	1.15 (.51)	1.21	1.13	1.44
ORGLAND	119.30 (468.27)	98.40	128.62	-0.00
LANDOWN	72.94 (161.21)	135.55	45.03	0.00
YRSCERT	5.52 (4.23)	4.62	5.92	-0.20
VARIETY	6.03 (5.47)	3.28	7.26	.11
VALUEADD	.53 (.96)	.31	.63	.66
TRANORB	1.69 (.46)	1	2	n/a
TOTYRS	14.84 (10.65)	15.63	14.50	0.07
ORGYRS	9.78 (7.44)	8.56	10.35	-0.10
AGE	47.15 (10.54)	47.36	47.06	0.07
EDUC	4.77 (1.40)	4.88	4.73	2.82
GENDER	1.79 (.45)	1.81	1.79	17.61

Notes: The first column reports the sample average and standard deviation for the entire sample of 396. The second and third columns report the sample averages for the transitioners and beginners subsamples, respectively. The fourth column is the t-statistics for the null hypotheses that the averages for the transitioners and beginners are identical.

Table 4: Preliminary tobit estimation results

Variables	**All Samples (1)**	**Transitioners (2)**	**Beginners (3)**
FINDMKTS	.13***	.23**	.08*
	(.043)	(.096)	(.046)
OBTACCS	-.07*	-.23**	-.04
	(.045)	(.114)	(.45)
NOTFINDP	.04	.093	.006
	(.038)	(.081)	(.041)
DISTANCE	-.10***	-.17**	-.05
	(.034)	(.078)	(.035)
RELPMT	.12***	.06	.13***
	(.038)	(.088)	(.040)
LACKCONS	-.10***	-.08	-.08**
	(.034)	(.077)	(.035)
LACKCERT	-.07	.05	-.03**
	(.049)	(.109)	(.037)
OVERSUP	.07**	.12	.03
	(.036)	(.083)	(.038)
TAMND	.0077***	-.000	.008***
	(.002)	(.006)	(.003)
TAMD	-.0013**	.001	-.002***
	(.001)	(.001)	(.001)
CONSTANT	0.76***	.94***	0.69***
	(.134)	(.289)	(.141)
Ancillary para	.69	.75	.60
N	360	111	249
Chi2 (d.o.f.)	44.42 (10)	17.14 (10)	29.04 (10)
p-value	0.00	0.07	0.001
Pesudo R2	0.07	0.09	0.06

Notes: ***: 1% significant; **: 5% significant; *: 10% significant. Standard errors are in parentheses. Dependent variable is the percentage of produce sold to indirect markets. There are 56 left-censored and 122 right censored observations in column (1). There are 9 left-censored and 65 right censored observations in column (2). There are 47 left-censored and 57 right censored observations in column (3).

Table 5: Baseline estimates of [4.6] and estimates accounted for heteroskedasticity, clustering and endogeneity

Variables	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
FINDMKTS	.085**	.041**	.085***	.041**	.081**	.039**
	(.038)		(.034)		(.036)	
OBTACCS	-.071*	-.035*	-.073**	-.036**	-.075**	-.036*
	(.040)		(.035)		(.036)	
NOTFINDP	.042	.021	.042	.020	.039	.019
	(.033)		(.036)		(.036)	
DISTANCE	-.084***	-.041***	-.083***	-.040***	-.081***	-.039***
	(.030)		(.027)		(.026)	
FAILCOMM	-.068*	-.033*	-.065	-.032	---	---
	(.038)		(.041)			
RELPMT	.13***	.063***	.131***	.063***	.100***	.048***
	(.039)		(.029)		(.022)	
LACKCONS	-.058**	-.028**	-.060**	-.029**	-.061**	-.029**
	(.031)		(.031)		(.030)	
LACKCERT	-.049	-.024	-.057*	-.028*	-.080***	-.038***
	(.046)		(.035)		(.032)	
OVERSUP	.102**	.050**	.088***	.043***	.083***	.040***
	(.044)		(.030)		(.030)	
TAMND	.004*	.002*	-.004***	.002***	.004***	.002***
	(.002)		(.001)		(.001)	
TAMD	-.001**	-.001**	-.001***	-.0005***	-.001***	-.0005***
	(.0005)		(.000)		(.000)	
BUSSTYPE	.142**	.069**	.134*	.065*	.126*	.061*
	(.070)		(.072)		(.068)	
ORGLAND	.000	.000	.000	.000	.000	.000
	(.000)		(.000)		(.000)	
LANDOWN	.000	.000	.000	.000	.000	.000
	(.000)		(.000)		(.000)	
YRSCERT	.0198*	.010*	.018	.009	.020	.009
	(.011)		(.013)		(.014)	
VARIETY	-.050***	-.024***	-.049***	-.024***	-.049***	-.024***
	(.007)		(.007)		(.007)	
VALUEADD	.072**	.035**	.069***	.033***	.061**	.029**
	(.037)		(.026)		(.025)	
TRANORB	-.331***	-.160***	-.341***	-.165***	-.353***	-.171***
	(.086)		(.092)		(.091)	
TOTYRS	-.000	-.000	-.001	-.0005	-.001	-.000
	(.004)		(.005)		(.005)	
ORGYRS	-.000	-.000	.000	.000	.000	-.000
	(.007)		(.001)		(.009)	
AGE	-.004	-.002	-.004	-.002	-.005	-.002
	(.004)		(.004)		(.004)	
EDUC	-.009	.004	-.004	-.002	-.005	.002
	(.026)		(.029)		(.029)	
GENDER	.153**	.074**	.154***	.075***	.165***	.080***
	(.077)		(.060)		(.059)	
CONSTANT	1.199***		1.286***		1.320***	
	(.314)		(.063)		(.319)	
(Wald) ch2	144 (23)		1171(23)		1030 (22)	
P-value	.0000		.0000		.0000	
R2 (log-l)	0.1984		-292		-.295	

Notes: ***: 1% significant; **: 5% significant; *: 10% significant. N = 360. Column (1) and (2) are Tobit estimates of coefficients and marginal effects for [4.6], respectively. Column (3) and (4) are coefficients estimates and marginal effects of specification [4.6] accounting for heteroskedasticity and clustering effect, respectively. Column (3) and (4) are coefficients estimates and marginal effects of specification [4.15] accounting for heteroskedasticity and clustering effect, respectively. Standard errors or robust standard errors are in parenthesis.

Table 6: Estimates of [4.15] by sample types

Variables	whole sample coef. (1)	whole sample mfx. (2)	transitioners coef. (3)	transitioners mfx. (4)	beginners coef. (5)	beginners mfx. (6)
FINDMKTS	.081** (.036)	.039**	.184*** (.074)	.070***	.041 (.039)	.022
OBTACCS	-.075** (.036)	-.036*	-.218* (.129)	-.083*	-.030 (.043)	-.016
NOTFINDP	.039 (.036)	.019	.095 (.059)	.036	.013 (.039)	.007
DISTANCE	-.081*** (.026)	-.039***	-.121 (.076)	-.046	-.073** (.033)	-.039***
RELPMT	.100*** (.022)	.048***	-.005 (.071)	-.002	.122*** (.033)	.066***
LACKCONS	-.061** (.030)	-.029**	-.061 (.087)	-.023	-.063** (.026)	-.034**
LACKCERT	-.080*** (.032)	-.038***	.073 (.087)	.028	-.115*** (.041)	-.063***
OVERSUP	.083*** (.030)	.040***	.120* (.068)	.046*	.065** (.034)	.035**
TAMND	.004*** (.001)	.002***	.002 (.004)	.001	.005*** (.001)	.003***
TAMD	-.001*** (.000)	-.0005***	.000 (.001)	.000	-.0016*** (.0001)	-.001***
BUSSTYPE	.126* (.068)	.061*	.350* (.189)	.133*	.047 (.063)	.026
ORGLAND	.000 (.000)	.000	-.000 (.000)	-.000	.000 (.000)	.000
LANDOWN	.000 (.000)	.000	-.000 (.000)	-.000	.001 (.001)	.001
YRSCERT	.020 (.014)	.009	.028 (.018)	.011	.022 (.014)	.012
VARIETY	-.049*** (.007)	-.024***	-.074*** (.020)	-.028***	-.041*** (.014)	-.022***
VALUEADD	.061** (.025)	.029**	.114 (.107)	.043	.054** (.025)	.029**
TRANORB	-.353*** (.091)	-.171***	---	---	---	---
TOTYRS	-.001 (.005)	-.000	-.005 (.009)	-.002	.001 (.008)	.001
ORGYRS	.000 (.009)	-.000	.000 (.018)	.000	-.001 (.001)	-.000
AGE	-.005 (.004)	-.002	.003 (.011)	.001	-.005 (.004)	-.003
EDUC	-.005 (.029)	.002	-.105 (.080)	-.040	.019 (.028)	.010
GENDER	.165*** (.059)	.080***	.180 (.172)	.068	.171*** (.062)	.093***
CONSTANT	1.320*** (.319)		.812 (.699)		.633* (.346)	
error term	.584		.645		.541	
N	360		111		249	
(Wald) ch2	1030 (22)		769(21)		241(21)	
P-value	.0000		.0000		.0000	
log-l	-295		-75		-207	

Notes: ***: 1% significant; **: 5% significant; *: 10% significant. All estimates account for heteroskedasticity and clustering effect by state of operation. Robust standard errors are in parenthesis. Column (1) and (2) are reproduced from last two columns in Table 5. Column (3) and (4) are coefficients estimates and marginal effects of specification [19] for the transitioners subsamples, respectively. Column (5) and (6) present the coefficients estimates and marginal effects of specification [19] for the beginners subsamples, respectively.

Table 7: Do Transaction Costs Matter: Tests Using Table 6

Null Hypothesis **H0: Joint significance of**	**whole sample** **chi2 (d.o.f)** **p-value** **reject?**	**transitioners** **chi2 (d.o.f)** **p-value** **reject?**	**beginners** **chi2 (d.o.f)** **p-value** **reject?**
(1) all transaction costs $\chi^2(10)$	112.83 .0000 reject	28.40 .0016 reject	65.37 .0000 reject
(2) all TC excluding infrastructure (TAMND, TAMD) $\chi^2(8)$	91.39 .0000 reject	19.45 .0126 fail to reject	42.08 .0000 reject
(3) (2)excluding market condition (LACKCONS, LACKCERT, OVERSUP) $\chi^2(5)$	77.88 .0000 reject	8.21 .1450 fail to reject	30.91 .0000 reject
(4) market condition and infrastructure $\chi^2(5)$	16.33 .0000 reject	8.61 .1256 fail to reject	27.00 .0001 reject
(5) market condition $\chi^2(3)$	18.76 .0001 reject	5.92 .1154 fail to reject	13.00 .0046 reject
(6) market infrastructure $\chi^2(2)$	18.67 .0001 reject	3.84 .1468 fail to reject	21.44 .0000 reject
(7) production characteristics $\chi^2(2)$	45.80 .0000 reject	15.12 .0005 reject	33.34 .0000 reject
(8) farm characteristics $\chi^2(5)$	9.46 .0919 fail to reject	3.76 .5846 fail to reject	12.65 .0268 fail to reject
(9) farmer characteristics $\chi^2(4)$	7.84 .0976 fail to reject	5.57 .2334 fail to reject	8.32 .0805 fail to reject

Notes: To reject or fail to reject the null is based upon 1% significance using estimates in Table 6. In (1), coefficients tested are FINDMKTS, OBTACCS, NOTFINDP, DISTANCE, RELPMT, LACKCONS, LACKCERT, OVERSUP, TAMND, TAMD. In (2), coefficients tested are FINDMKTS, OBTACCS, NOTFINDP, DISTANCE, RELPMT,LACKCONS, LACKCERT, OVERSUP. In (3), coefficients tested are FINDMKTS, OBTACCS, NOTFINDP, DISTANCE, RELPMT. In (4), coefficients tested are LACKCONS, LACKCERT, OVERSUP, TAMND, TAMD. In (5), coefficients tested are LACKCONS, LACKCERT, OVERSUP. In (6), coefficients tested are TAMND, TAMD. In (7), coefficients tested are VARIETY, VALUEADD. In (8), coefficients tested are TOTYRS, ORGYRS, AGE, EDUC, GENDER. In (9), coefficients tested are BUSSTYPE, ORGLAND, LANDOWN, YRSCERT.

Table 8: Estimates of [4.15] with simulated measurement errors

Variables	Whole Sample coef. (1)	Whole Sample mfx (2)	Transitioners coef. (3)	Transitioners mfx (4)	Beginners coef. (5)	Beginners mfx (6)
Panel (a): using simulated *uncorrelated* measurement errors						
FINDMKTS	.084***	.041***	.084	.031	.083***	.045***
	(.032)		(.073)		(.028)	
OBTACCS	-.096***	-.047***	-.173*	-.064*	-.076**	-.041**
	(.036)		(.102)		(.036)	
NOTFINDP	.032	.015	.066	.025	.011	.006
	(.033)		(.056)		(.036)	
DISTANCE	-.073***	-.035***	-.102	-.038	-.072*	-.040*
	(.035)		(.077)		(..040)	
RELPMT	.100***	.048***	.000	.000	.123***	.066***
	(.021)		(.051)		(.023)	
LACKCONS	-.033	-.016	.034	.013	-.051*	-.027*
	(.028)		(.077)		(.027)	
LACKCERT	-.057**	-.027***	.066	.025	-.079**	.039**
	(.029)		(.081)		(.037)	
OVERSUP	.089***	.043***	.148**	.055**	.075**	.040**
	(.030)		(.072)		(.034)	
TAMND	.005***	.002***	-.000	-.000	.006***	.003***
	(.001)		(.002)		(.001)	
TAMD	-.001***	-.000***	.001	.000	-.002***	-.001***
	(.000)		(.001)		(.000)	
BUSSTYPE	.135**	.065**	.321*	.119*	.072	.039
	(.070)		(.181)		(.060)	
VARIETY	-.049***	.024***	-.083***	-.031***	-.041***	-.022***
	(.008)		(.021)		(.0075)	
VALUEADD	.057***	.028***	.038	.014	.052***	.028***
	(.022)		(.088)		(.020)	
Panel (b): using simulated *correlated* measurement errors						
FINDMKTS						
OBTACCS						
NOTFINDP						
DISTANCE						
RELPMT						
LACKCONS						
LACKCERT						
OVERSUP						
TAMND						
TAMD						
BUSSTYPE						
VARIETY						
VALUEADD						

Notes: ***: 1% significant; **: 5% significant; *: 10% significant. Panel (a) is estimated with simulated uncorrelated measurement errors for three types of samples.

Table 9: Logit estimates of [4.19], [4.21] and [4.23]

Variables	Coefficient (1)	OR (2)	Coefficient (3)	OR (4)	Coefficient (5)	MFX (6)
FINDMKTS	.337 (.211)	1.402	.251* (.142)	1.285*	---	.058
OBTACCS	-.188 (.130)	.827	-.289 (.192)	0.749	---	-.060
DISTANCE	-.147 (.151)	.865	-.403*** (.132)	.668***	---	-.076
RELPMT	.387** (.163)	1.469**	.317*** (.107)	1.373**	---	.073
LACKCONS	-.050 (.168)	.950	-.308** (.134)	0.735**	---	-.056
LACKCERT	-.526*** (.145)	.594***	.185 (.180)	1.203	---	.012
OVERSUP	.347** (.182)	1.411**	.063 (.157)	1.065	.212*** (.066)	.022
TAMND	.011* (.007)	1.013*	.016* (.009)	1.016*	---	.005
TAMD	-.003** (.001)	.997**	-.004* (.002)	0.996*	---	-.004
BUSSTYPE	.808** (.417)	2.219**	.265 (.275)	1.304	.224* (.132)	
ORGLAND	.001 (.001)	1.001	.000 (.000)	1.000	-.0003*** (.000)	
LANDOWN	-.001 (.001)	.999	.004*** (.0015)	1.004***	.003* (.0017)	
YRSCERT	.115** (.195)	1.126**	.005 (.060)	1.005	.042** (.020)	
VARIETY	-.094*** (.028)	.911***	-.219*** (.052)	0.804***	-.079*** (.020)	-.091
VALUEADD	.580*** (.195)	1.781***	-.029 (.153)	.971	.019 (.051)	
TRANORB	-.950*** (.363)	-.385***	-.823** (.360)	0.439**	-.452 (.362)	-.432
TOTYRS	-.006 (.021)	.994	-.006 (.018)	0.994	.026** (.012)	
ORGYRS	.010 (.030)	1.009	.007 (.030)	1.007	-.037** (.017)	
AGE	-.044** (.019)	.956**	.009 (.016)	1.010	.010 (.014)	
EDUC	.048 (.165)	1.051	-.073 (.100)	0.930	.035 (.070)	
GENDER	.354 (.336)	1.413	.449 (.291)	1.567	.289* (.164)	
CONSTANT	3.363*** (1.148)	-	1.677 (1.335)	-	-.720 (.941)	
N	360		304		182	
Wald chi2	143.41 (21)		654.38(21)			
p-value	0.0000		0.0000			
Pseudo R2	0.1871		0.3075			

Notes: ***: 1% significant; **: 5% significant; *: 10% significant. All estimates account for heteroskedasticity and clustering effect of the state of operation. Robust standard errors are in parentheses. Column (1) and (2) present the coefficients estimates and odd-ratios of specification [4.19], respectively. Column (3) and (4) present the coefficients estimates and odd-ratios of specification [4.21], respectively. Column (5) lists the coefficients estimates of specification [4.23]. Column (6) lists the marginal effects of specifications [4.19],[4.21] and [4.23], with simulation size 1000 for each stage and covariates at sample means.

[i] Agriculture Marketing Service, United States Department of Agriculture, "The National Organic Program", 2002, available online at http://www.ams.usda.gov/nop/NOP/standards/FullText.pdf
[ii] By various sources: *Greene 2003, Nutrition Business Journal, November 2002*, Organic Trade Association at http://www.ota.org.
[iii] Furthermore, The United States is not a leader in the worldwide organic farming conversion. Farmers in 130 countries produce organically grown food and fiber on over 7 million hectares worldwide. Consumers worldwide spend $22 billion a year on organic products. (Worldwatch Institute, 2000). The United States ranked fourth in organic farmland, behind Australia (19 million acres), Argentina (6.9 million acres), and Italy (2.6 million acres). In terms of percentage of total farmland, the U.S. was much behind and was not among the top 10, which included Switzerland (9 percent), Austria (8.64 percent), Italy (6.76 percent), Sweden (5.2 percent), the Czech Republic (3.86 percent), and the United Kingdom (3.3 percent). (Greene 2003).
[iv] http://www.ers.usda.gov
[v] For example, while USDA tracks weekly prices at all market channels for a large number of conventionally grown commodities, the price information of the organics gathered by USDA is minimal.
[vi] Lohr et al. uses a matching approach and tests the likelihood of expansion for several market sectors based on the similarities between counties with and counties without organic markets based upon county-level data. They find that sales projections are overstated and that regional growth imbalance will continue.
[vii] Survey instruments are available from the author upon request. It is also available on http://www/ofrf.org.
[viii] Out of total 64 certification organizations identified. Several certification agencies like Quality Assurance International, Farm Verified Organic and Kauai Organic did not participate in releasing their member directory.
[ix] Data on farmers' markets is available at http://www.ams.usda.gov/farmersmarkets/map.htm.
[x] Data source is *National Organic Directory, Guide to Organic Information and Resource Worldwide*, 2001, published by Community Alliance with Family Farmers, http://www/caff.org.
[xi] See Hausman and Wise, 1985.
[xii] Note that the survey contains no separate transaction costs questions on direct markets and indirect markets. Answers to all transaction cost questions are individual farmers' experience of entering the overall organic markets, whether direct markets or indirect markets or both.
[xiii] The characteristics vector w that affect the transaction costs may overlap with, or be identical to, the characteristics vector z that affect the production quality distribution.
[xiv] The main results implied by condition [8] are similar to those in Goetz and Key et al.
[xv] TODO: list reference to show that, and give more details of direction and magnitudes of biases.
[xvi] TODO: comments on correcting for heteroskedasticity and clustering effects, and for endogeneity, may use hausman tests for both cases.
[xvii] See the discussion on the neglected heterogeneity issue in the binary response model in *Wooldridge (2002, p. 470)*
[xviii] To avoid computing a complex likelihood function when number of endogenous variables is great, Nelson and Olson provides a two-step estimator that is consistent yet inefficient.
[xix] TODO: hausman test, watch for negative covariance.
[xx] We intend to perform more experiments in our future work. We recognize that our current treatment to the measurement errors is rudimentary.
[xxi] TODO: test statistics go here.
[xxii] TODO: insert the figure
[xxiii] Coefficient estimate of YRSCERT in the reduced form [4.15] is *.020* but significant at *10%.* This does not conflict with the semi-reduced form estimates.