The Sample Selection Model: Application on The Farmers' Decision of Rice Acreage

Woraphon Yamaka[†,‡]**, Chaowana Phetcharat**[†,1]**,**

Nattamon Teerakul[†]**, and Thaya Navanugraha**[§]

[†]Faculty of Economics, Chiang Mai University, Chiang Mai 50200, Thailand

[‡]Excellence center in Econometrics, Faculty of Economics
Chiang Mai University, Chiang Mai 50200, Thailand
e-mail : woraphon.econ@gmail.com (Woraphon Yamaka)
chaowana@gmail.com (C. Phetcharat)
nattamon@econ.cmu.ac.th (N. Teerakul)

[§]Bureau of the Budget, Bangkok, Thailand
e-mail : nava_thaya@hotmail.com (T. Navanugraha)

Abstract : This study explores the estimation and inferences of a sample selection model with ordered-multinomial response outcomes. The model consists of two equations, namely, ordered response and regression equations. The estimation of parametric model is inconsistent when the distribution is unknown. To overcome this problem, we propose the primal generalized maximum entropy estimator to the model based on a constrained sample selection model with multinomial response outcomes. This method is also applied to examine farmers' decision on the rice acreage and perception of weather risk. The result of this study shows that the gender and age affect the farmers' subjective probability of the weather risk.

Keywords : Farmers' decision; generalized maximum entropy; rice planted area; weather events.
2010 Mathematics Subject Classification : 62P20; 91B84.

---

[1]Corresponding author.

1   Introduction

The goals of this study are to analyse farmers' choices of rice acreage, and to determine how the likelihood of unusual weather conditions influences the choices of rice planted area. The framework employed to evaluate the farmers' rice planted area decision follows the standard assumption that farmers are risk averse and concerning that an uncertain weather condition affects farm production costs and future returns. The rice planted area is therefore subjected to individual's past experience and likelihood of the unexpected weather events (i.e., flood and drought). The past experience of financial losses due to the disruptive weather could result in smaller rice acreage in the current year. Hence, the subjective probability of weather risk from farmers is considered when the farmers decide on their rice planted area. The rice acreage is determined in a two-stage process that considers farmers' given scores for a chance of unexpected weather events occurring on their farmlands, and then evaluates the alternative choices of rice planted area. A survey of rice producers in the northern region of Thailand provides the data for the analysis. To estimate the farmers' choice, the sample selection model with ordered outcomes of the Bellemare and Barrett[1] is applied. However, the problem of sample selection model often arises in many econometric studies. The maximum likelihood (ML) estimator may not be consistent when known distribution of the data generating process associated with the sample cannot be specified. Golan, Moretti, and Perloff [2] suggest that the conventional parametric and semiparametric estimations of the sample selection model are difficult to estimate the unknown parameters when sample sizes are small. For example, the full and limited information of the maximum likelihood as found in Heckman [3], and the semiparametric estimators of Ahn and Powell [4]. Furthermore, the choice of model assumptions for the conventional estimators is based on researcher's prior knowledge rather than the actual data generation process. This could lead to an inconsistent MLE when the assumptions are incorrectly specified.

In practice, the distribution of an error term is unknown, and none of the usual families of the distribution is a good fit to the residuals. The generalized maximum estimator (GME) requires a fewer number of assumptions for the error equation. The estimator uses the entropy-information measure [5] to recover the unknown probability distribution of underdetermined problems. The entropy refers to an amount of the uncertainty represented by a discrete probability distribution. Let x_j be a random variable with k possible outcomes. Each of the outcomes corresponds to the probability p_i which is unknown and unobserved. The sum of all outcomes probability equals to one, $\sum_i p_i = 1$. The information criterion (entropy) of Shannon's (1948) can be written as

$$H\left(P\right) = -\sum_i p_i \log p_i, \qquad (1.1)$$

where $p_i \log(p_i) = 0$ when $p_i = 0$. This entropy function measures the uncertainty of the distribution implied by P. The P reaches a maximum when $p_i = 1/N$.

Therefore, Generalized Maximum Entropy (GME) estimator by Golan, Judge, and Miller [6] was applied to our sample selection model. The model consists of two-stage regressions: 1) the farmers rating scores for the subjective probability of the unexpected weather events was classified by using the ordered choice model; and 2) the number of rice planted area was addressed with the regression model.

The contribution of this paper is threefold. First, we developed a semiparametric estimator for the sample selection model with ordered-multinomial response outcomes model. The estimator has its roots on the information theory, and it is based on the generalized maximum entropy (GME) approach of Golan, Judge, and Miller[6], Golan, Judge, and Perloff [7], Golan, Moretti, and Perloff [2]. Second, the validation of the GME estimator was performed. The method was done based on the data simulation. Third, the sample selection model in GME approach was used to determine the choice of the rice acreage.

The paper is organized as follows. Section 2 gives the explanation of the farm survey and data. Section 3 describes empirical methods of the sample selection model and GME estimator. Section 4 provides the validity of the GME estimator. The model application and conclusion of the study are presented in Sections 5 and 6, respectively.

2 Farm Survey and Data

A survey was conducted during October to November 2016 to identify the subjective probability of the expected weather events of rice producers and their planted area decision. The farm-level data was collected in three districts (San Kamphaeng, San Patong, and Hang Dong) of Chiang Mai Province, Thailand. The three districts were chosen because they are the major rice cultivation areas. One hundred twenty rice producers were randomly selected to take part in the survey. The survey questions include demographic information, farm structure, and weather risk experience. The questions on the ranking scores of the unexpected weather events are also inquired from the rice producers.

Table 1 reports mean and range values of farm characteristics for the sample selection model. Gender, education, and age of the respondents are explanatory variables in the ordered response equation. Household farm income and number of years in rice production are explanatory variables in the regression model. Most of the respondents are female with 39.1 years experience in rice production. The averages of respondent age and education are 61.5 and 2 years, respectively. The average farm income of the household is 44,458.83 Thai Baht/year (1 US dollar is approximately equal to 34 /Baht)

The farmers' rating scores indicating the chance of weather risk and their planted area decision are reported in Table 2. Most of the farmers, 42%, stated that their rice fields have moderate risk of the unexpected weather events, which may cause the farm future returns. Only 6% of the farmers thought the weather risk would have very low and low impact on their farmland. The mean of rice planted area is 3.44 acres (1 acre equals to 2.53 rai) which is two times lower than

Table 1: Mean and range values of farm characteristics by the regression model

Variable			
Ordered Response Equation			
Gender (1= male, 0 = female)	0.4583	0	1
Age (years)	61.5583	44	79
Education (Years of schooling)	2	1	3
Regression Equation			
Household farm income (baht/year)	44458.83	10000	150000
Number of years in rice production (years)	39.0667	7	55

Table 2: Descriptive statistic for the chance of weather risk and the rice planted area decision

Category	Frequency	Percentage
(Farmers rating scores for the chance of weather risk)		
Ordered response Equation		
Very low	1	0.09%
Low	7	5.83%
Moderate	50	41.67%
High	35	29.17%
Very high	27	22.50%
Regression Equation	**Mean**	**Min-Max**
Planted area of rice (rais)	8.7091	[3-23]

the maximum of the rice planted area.

About 78.3% of the households grow at least one variety of landrace rice and 21.7% grow other white rice. Only 9% of the households are willing to increase their cultivated areas. About 81.7% of the households insist to maintain the same cultivated area (Table 2).

3   Empirical Methods

This section develops the sample selection with ordered-multinomial response model which is subsequently implemented to the survey data in the next section. The idea behind the model is that because the unexpected weather events (flood and drought) from the farmers' prospective are partitioned into ordered categories. It is informative to distinguish the farmers' expectation on possible events rather than lump them together. Thus, the two equations, ordered outcomes and regression equations, were applied to the study. The ordered outcomes equation was used to classify the farmers' rating scores for the subjective probability of the unexpected weather events. The regression model was used to estimate the farmers' choice of rice planted area.

Table 3: Distribution of the upland rice growers

Category	Frequency	Proportion
Selection equation		
Grow landrace rice	260	0.7832
Not grow landrace rice (grow the white rice)	72	0.2168
Outcome Equation		
reduce the cultivated area	22	0.0877
maintain same size of the cultivated area	205	0.8167
increase the cultivated area	24	0.0956

3.1 The Sample Selection Model with Ordered-Multinomial Response Outcomes Model

The model consists of two equations, namely, ordered multinomial response and regression equations. Let's consider a model in which individuals $i(i = 1, ..., N)$ are sorted into $J + 1$ categories. The linear regression model under the ordered selection rule is written as follows:

$$s_i^* = \gamma G_i + v_i, \tag{3.1}$$

where s_i^* is a ordinal response variable of an individual i, which takes a value of $0, ..., J$. That is

$$s_i = \begin{cases} 1 \text{ if } -\infty s_i^* \le \mu_1 \\ 2 \text{ if } \mu_1 s_i^* \le \mu_2 \\ 3 \text{ if } \mu_2 s_i^* \le \mu_3 \\ \vdots \\ J \text{ if } \mu_J s_i^* \le \infty \end{cases} \tag{3.2}$$

where G_i is the vector of explanatory variables of the ordered-multinomial response equation. γ is the vector of estimated parameters, v_i is the sequence of unobservable random errors, and $\mu_1, ..., \mu_J$ are the unknown cut-offs which satisfy $\mu_1 < \mu_2 < \cdots < \mu_J$. We assume that the independent variables G_i and the categorical variable s_i are observed, but the latent selection variable s_i^* is unobserved.

For regression equation, we observe a dependent variable y_i that is a linear function of the vector of explanatory variables X_i. The regression equation can be specified as

$$y_i = X\beta_i + \varepsilon_i \tag{3.3}$$

$$y_i = \begin{cases} \beta_1 X_{1i} + \varepsilon_{1i} & \text{if} \quad s_i = 1, \\ \beta_2 X_{2i} + \varepsilon_{2i} & \text{if} \quad s_i = 2, \\ \quad\quad \vdots \\ \beta_J X_{Ji} + \varepsilon_{Ji} & \text{if} \quad s_i = J, \end{cases} \tag{3.4}$$

where β is the vector of estimated parameters depending on the category s_i, and ε_i is a random error term for the regression equation. Technically, no assumptions for the distribution of the errors ε_i and v_i are made in this study.

3.2 The Formulation of the Sample Selection with Ordered-Multinomial Response Model

The GME approach was used to estimate the sample selection with ordered-multinomial response data model. First, we start by providing the background of the generalized maximum entropy approach and then develop the GME for our study model.

In GME, the parameter vectors of γ, β, μ_i, v_i, and ε_i are decomposed into a set of categories M, support points z_k, and probability weights p_k. The support points are commonly provided based on the researcher prior information. In the ordered-multinomial response equation, each parameter of the explanatory variables (γ) has M categories (where $M > 2$), support points r_k, and probability weights p_k (k =1, ..., K). Each parameter of the explanatory variables in regression equation (β) has $M > 2$ categories, support points r_k, and probability weights q_k (k =1, ..., K). Likewise, the parameter of the each cut-off value, μ_j, has $M > 2$ categories, support points c_j, and probability weights l_j (j=1,,J). The unobservable random errors v_i and ε_i also have $M > 2$. There are u_i support points and ω_i probability weights for each error v_i while each error ε_i has support points v_i and probability weights w_i. Intuitively, each parameter and error is equal to the product of a support point and its associated probability weight, summed over all support points. In the model, we specify the support space for γ, β, μ_j, v_i, and ε_i to be both negative and positive possible values. Thus, the parameters of Eqs. 3.1–3.2 are reparameterized as

$$\gamma = Zp = \begin{bmatrix} z'_1 & 0 & \cdots & 0 \\ 0 & z'_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & z'_K \end{bmatrix} \cdot \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_K \end{bmatrix}, \tag{3.5}$$

$$\beta = Rq = \begin{bmatrix} r'_1 & 0 & \cdots & 0 \\ 0 & r'_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & r'_K \end{bmatrix} \cdot \begin{bmatrix} q_1 \\ q_2 \\ \vdots \\ q_K \end{bmatrix}, \tag{3.6}$$

$$\mu = CL = \begin{bmatrix} c'_1 & 0 & \cdots & 0 \\ 0 & c'_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & c'_J \end{bmatrix} \cdot \begin{bmatrix} l_1 \\ l_2 \\ \vdots \\ l_J \end{bmatrix}, \tag{3.7}$$

$$
v = U\omega = \begin{bmatrix} u'_1 & 0 & \cdots & 0 \\ 0 & u'_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & u'_N \end{bmatrix} \cdot \begin{bmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_N \end{bmatrix}, \tag{3.8}
$$

$$
\varepsilon = Vw = \begin{bmatrix} v'_1 & 0 & \cdots & 0 \\ 0 & v'_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & v'_N \end{bmatrix} \cdot \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \end{bmatrix}, \tag{3.9}
$$

Combining Equations 3.5-3.9, the generalized sample selection with ordered-multinomial response data model to be estimated is

$$
s_i = \begin{cases} 1 \ \text{if} \ \ \infty ZpG'_1 + U_1\omega_1 \leq c'_1 l_1 \\ 2 \ \text{if} \ c'_1 l_1 ZpG'_2 + U_2\omega_2 \leq c'_2 l_2 \\ 3 \ \text{if} \ c'_2 l_2 ZpG'_3 + U_3\omega_3 \leq c'_3 l_3 \\ \qquad \vdots \ \text{if} \\ J \ \text{if} \ c'_J l_J ZpG'_J + U_J\omega_J \leq \infty \end{cases} \tag{3.10}
$$

$$
y_i = \begin{cases} R_1 q_1 X_{1i} + V_1 w_1 \quad \text{if} \quad s_i = 1 \\ R_2 q_2 X_{2i} + V_2 w_2 \quad \text{if} \quad s_i = 2 \\ \qquad\qquad \vdots \\ R_J q_J X_{Ji} + V_J w_J \quad \text{if} \quad s_i = J \end{cases} \tag{3.11}
$$

3.3   Maximum Entropy Generalized Maximum Entropy Estimators

After reparameterizing the model in Eqs. 3.10–3.11, the GME objective function can be expressed as

$$
\max_{p,q,l,\omega,w} (p, q, l, \omega, w) = -p' \log p - q' \log q - l' \log l - \omega' \log \omega - w' \log w \tag{3.12}
$$

Subject to:

$$
s_i = \begin{cases} 1 \ \text{if} \ \ \infty ZpG'_1 + U_1\omega_1 \leq c'_1 l_1 \\ 2 \ \text{if} \ c'_1 l_1 ZpG'_2 + U_2\omega_2 \leq c'_2 l_2 \\ 3 \ \text{if} \ c'_2 l_2 ZpG'_3 + U_3\omega_3 \leq c'_3 l_3 \\ \qquad \vdots \ \text{if} \\ J \ \text{if} \ c'_J l_J ZpG'_J + U_J\omega_J \leq \infty \end{cases} \tag{3.13}
$$

$$
y_i = \begin{cases} R_1 q_1 X_{1i} + V_1 w_1 \quad \text{if} \quad s_i = 1, \\ R_2 q_2 X_{2i} + V_2 w_2 \quad \text{if} \quad s_i = 2, \\ \qquad\qquad \vdots \\ R_J q_J X_{Ji} + V_J w_J \quad \text{if} \quad s_i = J, \end{cases} \tag{3.14}
$$

and adding up constraints

$$1'p = 1 \quad k = 1, ..., K_1, \tag{3.15}$$

$$1'q = 1 \quad k = 1, ..., K_2, \tag{3.16}$$

$$1'l_j = 1 \quad j = 1, ..., J, \tag{3.17}$$

$$1'\omega_j = 1 \quad i = 1, ..., N, \quad j = 1, ..., J, \tag{3.18}$$

$$1'w_j = 1 \quad i = 1, ..., N_j \quad , j = 1, ..., J \tag{3.19}$$

The corresponding Lagrangian is

$$
\begin{aligned}
L = {} & H(p, q, l, \omega, w) + \lambda_1^{(1)} (c'_1 m_1 - ZpG'_1 - U_1\omega_1) + ... + \lambda_J^{(1)}(c'_J l_J - ZpG'_1 - U_J\omega_J) \\
& + \lambda_1^{(2)} (c'_1 l_1 - ZpG'_2 - U_2\omega_2) + ... + \lambda_{J-1}^{(2)} (c'_{J-1} l_{J-1} - ZpG'_J - U_J\omega_J) \\
& + \theta_1(y_{i1} - R_1 q_1 x_{i1} - V_1 w_1) + ... + \theta_J(y_{iJ} - R_J q_J x_{iJ} + V_J w_J) + \phi_1(1 - 1'p) \\
& + \phi_2(1 - 1'q) + \phi_3(1 - 1'l_j) + \phi_4(1 - 1'\omega_j) + \phi_5(1 - 1'w_j)
\end{aligned}
\tag{3.20}
$$

where λ , θ , ϕ are the vectors of Lagrangian multiplier. Taking the gradient of L to derive the first-order condition, resulting in the solutions as

$$\widehat{p}_{km} = \frac{\exp(-ZG'_{1k}\tilde{\lambda}_1 - ... - ZG'_{Jk}\tilde{\lambda}_J)}{\sum\limits_m \exp(-ZG'_{1k}\tilde{\lambda}_1 - ... - ZG'_{Jk}\tilde{\lambda}_J)}, \quad k = 1, ..., K_1 \tag{3.21}$$

$$\widehat{q}_{jkm} = \frac{\exp(-r_{jkm}X'_{jk}\tilde{\theta}_j)}{\sum\limits_m \exp(-r_{jkm}X'_{jk}\tilde{\theta}_j)}, \quad j = 1, ..., J \ , k = 1, ..., K_2 \tag{3.22}$$

$$\widehat{l}_{jm} = \frac{\exp(-\widehat{\lambda}^{(1)} c_{jm} - \widehat{\lambda}^{(2)} c_{jm})}{\sum\limits_m \exp(-\widehat{\lambda}^{(1)} c_{jm} - \widehat{\lambda}^{(2)} c_{jm})}, \tag{3.23}$$

$$\widehat{\omega}_{im} = \frac{\exp(-U_1\widehat{\lambda}_1 - ... - U_J\widehat{\lambda}_J)}{\sum\limits_m \exp(-U_1\widehat{\lambda}_1 - ... - U_J\widehat{\lambda}_J)}, \tag{3.24}$$

$$\widehat{w}_{jim} = \frac{\exp(-\widehat{\theta}_j V_j)}{\sum\limits_m \exp(-\widehat{\theta}_j V_j)}, \tag{3.25}$$

The optimal solution of the equations 3.21–3.25 yields the point estimates

$$\widehat{\gamma}_k = z'_{km}\widehat{p}_{km}, \quad k = 1, ..., K_1, \tag{3.26}$$

$$\widehat{\beta}_{jk} = \sum_m r'_{jkm}\widehat{q}_{km} \ , \quad j = 1, ..., J \ , k = 1, ..., K_2, \tag{3.27}$$

$$\mu_j = \sum_m c_{jm}l_{jm}, \quad j = 1, ..., J, \tag{3.28}$$

$$v_i = \sum_m u'_{im}\omega_{im}, \quad i = 1, ..., N, \tag{3.29}$$

$$\varepsilon_{jm} = \sum_m \mathrm{v}_{jim}\widehat{w}_{jim}, \quad j = 1, ..., J \ , i = 1, ..., N_j. \tag{3.30}$$

3.4   Validation of the Generalized Maximum Entropy (GME) Estimator

To examine the accuracy of the GME estimator for the sample selection with ordered-multinomial response model, the simulation study is conducted.

3.4.1   Data-Generating Process

G_i is random variable which is independently generated under the standard uniform distribution. The error terms v_i and ε_i are generated under the standard bivariate normal distribution with the correlation value of 0.5. The selection process is

$$s_i^* = \gamma G_i + v_i, \tag{3.31}$$

$$s_i = \begin{cases} 1 \text{ if } -\infty < \gamma G_i' + v_i \leq 0 \\ 2 \text{ if } 0 < \gamma G_i' + v_i \leq 1 \\ 3 \text{ if } 1 < \gamma G_i' + v_i \leq \infty \end{cases} \tag{3.32}$$

The dependent variable y_i is defined by

$$y_i = \begin{cases} \beta_1 X_i + \varepsilon_i \text{ if } s_1 = 1 \\ \beta_2 X_i + \varepsilon_i \text{ if } s_1 = 2 \\ \beta_3 X_i + \varepsilon_i \text{ if } s_1 = 3 \end{cases} \tag{3.33}$$

where X_i is random variable generated under the uniform distribution. The sample sizes of one hundred and two hundred ($N = 100$ and $N = 200$) are considered for generating the random variables. The values for parameters γ, β_1, β_2 and in Eqs. (3.32–3.33) are 0.5, 1, 2, and 3, respectively. The 100 repetitions are made for each simulation.

In this simulation study, we fix $M = 5$ for the number of support points. Since the true value of the estimated parameters is known, we can determine the upper and lower bounds of the parameters and errors covering the true value. Thus, we set $z = r = [-5, -2.5, 0, 2.5, 5]$, and $c = [-2, -10, 1, 2]$. For the error supports, we specified the values using three sigma rule of Pukelsheim [8] and thereby setting $u = v = [-3\sigma^2, -1.5\sigma^2, 0, 1.5\sigma^2, 3\sigma^2]$.

3.4.2   Results of the Simulation Study

Tables 3 displays the summary statistics of the parameters obtained from the 100 replications. Columns denote the true values, mean, and standard deviation of the estimated parameters. The simulation results show that the mean estimates are close to the true values. The standard error gets smaller as the sample size increases. From these results we can infer that the GME is reliable and accurate estimator for our model. Moreover, we also compare the performance of GME with two-step estimation of Greene [9], we find that the GME performs better than two-step estimation, when the data is small ($N = 100$).

<div align="center">Table 4: Simulation results</div>

Par.	TRUE	GME		Two-step estimation	
		N=100	N=200	N=100	N=200
γ	0.5	0.3246	0.4668	1.134	0.179
		(0.1911)	(-0.0125)	(-0.6257)	(-0.4421)
β_1	1	1.2095	1.3787	1.0898	1.3787
		(-0.2637)	(-0.1302)	(-0.2184)	(-0.1302)
β_2	2	2.2846	2.3511	2.166	2.3511
		(-0.2389)	(-0.1328)	(-0.211)	(-0.1328)
β_3	3	3.0815	3.1941	2.9492	3.1941
		(-0.1929)	(-0.1374)	(-0.1677)	(-0.1374)
μ_0	0	0.1615	0.053	0.1615	-0.253
		(-0.1515)	(-0.0652)	(-0.3548)	(-0.2652)
μ_1	1	1.1161	0.9384	1.358	0.9384
		(-0.3010)	(-0.2733)	(-0.3807)	(-0.2733)

Source: Calculation

4   Results of Application Study

In this empirical study, we also fix $M = 5$ for the number of support points where $z = r = c = [-2 - 1, 0, 1, 2]$. Furthermore, we have our expectation that the range of the estimated parameters is large, and the bound of the coefficients would be located between $[-2, 2]$ and $[-4, 4]$ for cutoffs parameters. This is due to we have only 5 ordered choice scores. For the error supports, we specify the values using three sigma rule of Pukelsheim [8] and set $u = v = [-3, -1.5, 0, 1.5, 3]$.

Table 4 reports the estimated coefficients for sample selection with ordered-multinomial response model. The ordered response and regression equations are obtained from this model. The gender and age significantly affect the farmers' subjective probability of the weather risk. Female–rice farmers and older ages give a higher score for the probability of the weather risk occurrence.

For the regression equation of the rice planted area decision, the farmers in all categories of the weather risk prospective (s_i) are affected significantly by the household average farm income (1% level). Higher farm income household earned per year increases the rice planted area. This implies that the farmers spend more

Table 5: Estimation results

Variable		Coefficient	Standard error
Ordered response Equation			
Gender		0.0190*	0.1998
Age		0.0230*	0.017
Education		-0.8447	0.7891
Regression Equation in each ordered response			
$s=1$	Income	0.0003***	0.0001
	Experience	0.00001	0.0001
$s=2$	Income	0.0003***	0.0001
	Experience	0.3743***	0.0049
$s=3$	Income	0.00001***	0.00002
	Experience	0.1691***	0.0248
$s=4$	Income	0.0001***	0.00001
	Experience	0.0933**	0.0287
$s=5$	Income	0.00003*	0.00001
	Experience	0.1647***	0.0184
μ_1		-2.7294	2.3753
μ_2		-1.8031	2.3386
μ_3		-0.2965	2.3327
μ_4		0.509	2.3332

Note: ***, **, and * represent significance at 1%, 5%, and 10% level, respectively

time on farming regardless of their higher future returns. The number of years in rice production also has positive and significant impact on the farmers' decision. The longer time farmers have been in rice production, the larger size of the rice acreage. Only those farmers who expect the weather risk is unlikely to occur on their farmland ($s = 1$) are not affected by the farming experience.

5   Conclusions

The problem of sample selection model often arises in many econometric studies. The maximum likelihood (ML) estimator is not consistent when the distribution of the data generating process associated with the sample is not known. In addition, our data sets are relatively small and need a consistent estimator that recovers these problems and converges to an optimal solution. Thus, we employ the generalized maximum entropy(GME) estimator to estimate all unknown parameters of the sample selection with ordered-multinomial response outcomes model. We also conduct the simulation study to examine the accuracy of the model and find that our model performs well in the simulation study.

From the application study, our results shed some light on how farmers make a decision on the rice planted area regardless of the unusual weather events. The gender and age affect the farmers' subjective probability of the weather risk. The female farmers and older ages have a higher prospective of the unexpected weather events and are more concern about the weather risk. The annual household in-

come from farming also increases the rice planted area. The farmers' decision is positively impacted by the farmers' experience. The longer time farmers have been in rice production, the larger size of the rice acreage.

References

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

[2] A. Golan, E. Moretti, J. M. Perloff, A small-sample estimator for the sample-selection model, Econometric Reviews 23 (1) (2004) 71–91.

[3] J. Heckman, Sample selection bias as a specification error, Econometrica 47 (1) (1979) 153-161.

[4] H. Ahn, J. L. Powell, Semiparametric Estimation of Censored Selection Models with a Nonparametric Selection Mechanism, Journal of Econometrics 58 (1993) 3–29.

[5] C. E. Shannon, A Mathematical Theory of Communication, Bell System Technical Journal 27 (1948) 379–423.

[6] A. Golan, G. G. Judge, D. Miller, Maximum Entropy Econometrics: Robust Estimation with Limited Data, Wiley, New York, 1996.

[7] A. Golan, G. Judge, J. Perloff, Estimation and inference with censored and ordered multinomial response data, Journal of Econometrics 79 (1) (1997) 23–51.

[8] F. Pukelsheim, The three sigma rule, The American Statistician 48 (2) (1994) 88-91.

[9] W. H. Greene, Econometric Analysis, Fifth Edition, Prentice Hall, 2003