PRACE NAUKOWE AKADEMII EKONOMICZNEJ WE WROCŁAWIU
Nr 1169 2007
TAKSONOMIA 14
Klasyfikacja i analiza danych – teoria i zastosowania

Hanna Dudek
Szkoła Główna Gospodarstwa Wiejskiego w Warszawie

AN IDENTIFICATION OF FARMERS' HOUSEHOLDS IN DANGER OF POVERTY ON THE GROUND OF ORDERED LOGIT MODEL

1. Introduction

Farmers are group of society that has been particularly affected by the economic transformation in Poland. Uncertain markets and prices for agricultural goods have worsened conditions of their households. It is easy to talk about rural poverty as an evil which should be eliminated. However, it is, not so simple to identify the households in danger. In that paper we present the results of attempt of their identification by using ordered logit models.

The poverty itself is usually referred to such situations where there is scarcity of economic goods, standard of living and social services which are available to the majority of people. The poverty is a multidimensional phenomenon covering such important areas as: employment, health, education, financial poverty. In the study we focus on the last one.

Poverty can be approached from either subjective or objective perspective. In objective approach the status of households can be verified by documentary evidence. Subjective perceptions base on question to households about their perceived situation. The paper deals with both approach.

When estimating monetary measures of poverty, one may have a choice between using disposable income or total expenditures as the indicator of well-being. Some analysts argue that expenditures may better show poverty than income since the last may be erratic and fluctuate during the year. It is important especially in the case of farmers' households. The fluctuation of monthly income in that case is due to character of agricultural production. Moreover inaccuracy in registration of income is caused by existence of large informal sectors. On the other hand, some of

the issues involved in total expenditure refer to purchase of durable good and repair service. Therefore there are some pros and cons for each approach. In objective approach we decide to use total monthly equivalent expenditures of households as welfare measure applying OECD 70/50 equivalence scale.

2. Data

The empirical analyses are based on the Household Budget Survey carried out by the Central Statistical Office. It encompassed 1241 farmers' households in 2003. The data from household budget survey covered information on household incomes, expenditures and on household demographic and socio-economic attributes. In this paper those households are taken into account whose exclusive or main source of livelihood was income from used private farm in agriculture. Additional sources of maintenance for this group of households may include old age pension, other types of pension or any other unearned income, self-employments outside a private farm in agriculture, or free profession. The income gained from additional sources is lower than from the private farm in agriculture [*Hausehold...* 2004].

Analysis of subjective poverty in that study is based on question: "It is possible in your household to meet two ends with the present income?" Possible answers are: very poor, rather poor, insufficiently, scarcely enough, good very good.

There are only one household assessing its situation as "very good" - we delete it. Due to a small number of households (13) declaring their situation as good, we join them with those estimating their income position as scarcely enough.

In analysis of objective poverty we fix relative poverty line at 60% of median of equivalent total expenditures. The median of equivalent expenditures in 2003 was:

- for all households in Poland - 773 zl, thus 60% this value gives 464 zl,
- for farmers' households – 585 zl, thus 60% this value gives 351 zl.

In that case we consider five categories as presented in table 1.

Table 1. Number of households with distinct levels of welfare

Subjective Poverty			Objective Poverty		
Income Status	category	number of households	equivalent expenditures	category	number of households
Very poor	$j = 1$	247	below 351	$j = 1$	174
Poor	$j = 2$	338	[351; 464)	$j = 2$	226
Insufficient	$j = 3$	542	[464; 585)	$j = 3$	221
Scarcely enough or good	$j = 4$	114	[585; 773)	$j = 4$	273
-	-	-	at least 773	$j = 5$	346
Total	all	1240	total	all	1240

Source: own calculations based on Household Budget Survey data.

In order to indicate the groups of high risk of poverty we apply ordered models.

3. Models for Ordinal Outcomes

Models for ordinal outcomes take into account the ordered nature of the response. In these models ordinal variable is often coded as consecutive integers from one to the number of categories. Such variable clearly order the categories but absolute distances between categories are unknown, therefore the linear regression models analysis is inappropriate. It is much better to use models designed for ordinal outcomes. The starting point in such case is usually an model with latent variable [Long, Freese 2001]:

$$y_i^* = \mathbf{x}_i^T \boldsymbol{\beta} + \varepsilon_i, \; i = 1, 2, ..., n,$$

where:

– $\boldsymbol{\beta}$ – a column vector of parameters $\beta_1, ...\beta_k$,

– $\mathbf{x}_i$ – a row vector representing the characteristics of individual i,

– ε_i –random error,

– n – number of individuals,

– y^* – the latent variable, which represents the response, if it could be measured accurately on the continuous scale. Let us assume a set of cut-points $\delta_0, \delta_1, \cdots \delta_m$, such that $-\infty = \delta_0 < \delta_1 < \cdots \delta_m = \infty$, that divide $(-\infty; \infty)$ into m intervals. The relationship between the latent variable and the realized outcome is: $y_i = j$ if and only if $\delta_{j-1} < y_i^* \le \delta_j$ [1], $i = 1,2, ..., n, j = 1, ..., m$. As $\delta_{j-1} < y_i^* \le \delta_j$ and $y_i^* = \mathbf{x}_i^T \boldsymbol{\beta} + \varepsilon_i$, so using some algebra we have: $\delta_{j-1} - \mathbf{x}_i \boldsymbol{\beta} < \varepsilon_i \le \delta_j - \mathbf{x}_i \boldsymbol{\beta}$. It leads to the following probabilities of each outcome:

$P(y_i = j | \mathbf{x}_i) = F(\delta_j - \mathbf{x}_i \boldsymbol{\beta}) - F(\delta_{j-1} - \mathbf{x}_i \boldsymbol{\beta})$, where F – cdf of iid error terms ε_i.

In applications usually following models are used [Borooah 2001; *Stata*... 2005]:

- ordered logit model with $F(z) = \Lambda(z) = \dfrac{1}{1 + \exp(-z)}$,
- ordered probit model with $F(z) = \Phi(z) = \int_{-\infty}^{z} \dfrac{1}{\sqrt{2\pi}} \exp\left(-\dfrac{t^2}{2}\right) dt$.

In this paper we focus on the first one. The slope parameters $\beta_1, \beta_2, \ldots, \beta_k$ have no intuitive interpretation, but we can examine their signs to determine directions of changes of probabilities. Let us compute the marginal effects in the probability:

---

[1] The $\delta_0, \delta_1, \ldots, \delta_m$ are unknown parameters to be estimated with $\beta_1, \beta_2, \ldots, \beta_k$.

$$\frac{\partial P(y_i = j|\mathbf{x})}{\partial x_l} = -\beta_l \left\{ \frac{d\Lambda(\delta_j - \mathbf{x}\boldsymbol{\beta})}{dz} - \frac{d\Lambda(\delta_{j-1} - \mathbf{x}\boldsymbol{\beta})}{dz} \Big|_{z=\mathbf{x}\boldsymbol{\beta}} \right\}.$$

The term in braces can be positive or negative, so one must be very careful in interpreting the slope parameters $\beta_1, \beta_2, \ldots, \beta_k$ in the model [Greene 2000]. Only the signs of the changes in $P(y_i = 1|\mathbf{x})$ and $P(y_i = m|\mathbf{x})$ are unambiguous. The marginal effects of the regressor x_l on the probabilities $P(y_i = 1|\mathbf{x})$ are: $\frac{\partial P(y_i = 1|\mathbf{x})}{\partial x_l} = -\beta_l \left(\Lambda(\delta_1 - \mathbf{x}_i\boldsymbol{\beta})(1 - \Lambda(\delta_1 - \mathbf{x}_i\boldsymbol{\beta})) \right)$, l = 1, 2, ..., k. Since $\Lambda(1-\Lambda) \geq 0$, then the derivative of $P(y_i = 1|\mathbf{x}_i)$ has the opposite sign from β_l. By similar logic, as $\frac{\partial P(y_i = m|\mathbf{x})}{\partial x_l} = \beta_l \left(\Lambda(\delta_{m-1} - \mathbf{x}_i\boldsymbol{\beta})(1 - \Lambda(\delta_{m-1} - \mathbf{x}_i\boldsymbol{\beta})) \right)$, therefore the change in $P(y_i = m|\mathbf{x})$ must have the same sign as β_l.

The above approach for analyzing marginal effects is only appropriate when the regressor is continuous. The effects of a dummy variable should be analyzed by comparing the probabilities computed when the dummy variable takes respectively one and zero value, remaining the values of other variables unchanged.

4. Estimation and Hypothesis Testing

The parameters of ordered logit model we estimate by maximum likelihood method (ML). We consider a following likelihood function:

$$L(\boldsymbol{\beta}, \boldsymbol{\delta}|\mathbf{x}) = \prod_{i=1}^{n} \prod_{j=1}^{m} P(y_i = j|\mathbf{x}_i)^{d_{ij}} =$$

$$\prod_{i=1}^{n} \prod_{j=1}^{m} \left[\frac{1}{1 + \exp(\mathbf{x}_i\boldsymbol{\beta} - \delta_j)} - \frac{1}{1 + \exp(\mathbf{x}_i\boldsymbol{\beta} - \delta_{j-1})} \right]^{d_{ij}},$$

where: $d_{ij} = 1$ if $y_i = j$ and $d_{ij} = 0$ otherwise.

The values of the parameters that maximize this function are the maximum likelihood estimates. We find them using Stata Statistical Software.

Ordered logit model approach assumes parameters $\beta_1, \beta_2, \ldots, \beta_k$ to be the same for all categories. Some researchers [Long, Freese 2001; *Stata...* 2005] maintain that validity of this assumption should be tested by comparing the likelihood value obtained by fitting the ordered logit model with unordered one (this model allows

those parameters to be different between the outcomes $j = 1, 2, \ldots, m$). If $\ln\hat{L}_o$ and $\ln\hat{L}_m$ are respectively the log-likelihood values from the ordered model and multinomial (unordered) one, then we can compute $LR_{om} = -2(\ln\hat{L}_o - \ln\hat{L}_m)$ and compare it to $\chi^2(k(m-2),\alpha)$. The result of this test should be interpreted carefully, because ordered logit model is not nested within multinomial one. A large values of LR_{om} can be however taken as evidence of poorness of fit.

For testing hypothesis about parameters, many procedures are available. The simplest method for a single restriction uses the standard normal table for critical points. We can also test a hypothesis that involves more than one parameter. For a test of the null hypothesis that all slope parameters are zero, the likelihood ratio test statistic is calculated as:

$$LR = -2(\ln\hat{L}_{Intercept} - \ln\hat{L}_{Full}),$$

where $\ln\hat{L}_{Intercept}$ denotes the value of the restricted log-likelihood when all slope coefficients are zero and $\ln\hat{L}_{Full}$ – the log-likelihood value for full model (without any restrictions imposed on parameters). Approximate critical values are obtained from the chi-square distribution with degrees of freedom equal to the number of slope parameters.

There are a wide variety of measures of the goodness of fit often called pseudo-R-square statistics. Most often cited is measure based on likelihood ratio (also known as McFadden[2] R-Squared): $pseudo - R^2 = 1 - \dfrac{\ln\hat{L}_{Full}}{\ln\hat{L}_{Intercept}}$.

Because this statistic does not mean what R-square means in an conventional regression, some researchers suggest interpreting this statistic with great caution.

5. Results

At the first stage we considered potential variables that can explain differentiation in financial status of households. We took into account many attributes of the household head and characteristics referring to whole household [Panek 1991]. In the next stage we followed in accordance with statistical criteria recommended in literature [Gruszczyński 2002]. Variables regarded in ordered logit models were:

- X1 – equals 1, if household consist only with couple without children, 0 – otherwise,
- X2 – and area of farm in hectares,
- X3 – number of dependants in a household,

[2] Daniel McFadden is a winner (jointly with James Heckman) of the 2000 Nobel Prize in Economic Sciences "for his development of theory and methods for analyzing discrete choice".

- X4 equals 1, if household's head has at least secondary education, 0 – otherwise,
- X5 equals 1, if there is hot running water available in a household, 0 – otherwise,
- X6 equals 1, if there is a stationary phone in a household, 0 – otherwise,
- X7 – number of rooms in residence,
- X8 equals 1, if household is in mazowieckie province (voivodeship), 0 – otherwise,
- X9 equals 1, if household is in podlaskie province, 0 – otherwise,
- X10 equals 1, if household is in warmińsko-mazurskie province, 0 – otherwise,
- X11 –number of sources of livelihood,
- X12 – equals 1 household's head is woman, 0 – man,
- X13 – number of jobless due to illness,
- X9' equals 1, if household is in podkarpackie province, 0 – otherwise,
- X10' equals 1, if household is in zachodniopomorskie province, 0 – otherwise.

The obtained results are presented in table 2.

Table 2. Ordered logit models estimation results

Subjective poverty			Objective poverty		
Explanatory variable	parameter estimate	standard error	explanatory variable	parameter estimate	standard error
X1	0,63	0,25	X1	0,66	0,25
X2	0,02	0,01	X2	0,02	0,004
X3	-0,16	0,04	X3	-0,48	0,04
X4	0,43	0,15	X4	0,51	0,15
X5	0,73	0,14	X5	0,30	0,14
X6	0,57	0,13	X6	1,12	0,13
X7	0,14	0,04	X7	0,23	0,04
X8	0,53	0,14	X8	0,25	0,13
X9	-0,40	0,20	X9'	-0,76	0,35
X10	0,90	0,28	X10'	-1,00	0,40
X11	0,19	0,05	X11	-0,28	0,05
X12	-0,44	0,16	-	-	-
X13	-0,70	0,23	-	-	-
Pseudo R^2 = 0,09, Log L = -1421,41			Pseudo R^2 = 0,10, Log L = -1769,82		
LR = 268,19, $LR_{om} = -2(\ln \hat{L}_o - \ln \hat{L}_m) = 24,14$			LR = 381,21, $LR_{om} = -2(\ln \hat{L}_o - \ln \hat{L}_m) = 33,78$		

Source: own calculations obtained by using Stata Software.

First, tests were conducted to assess the statistical significance of the results. For both models values of LR_{om} statistic don't exceed critical values, so we can state that application of ordered logit models is justifiable. Values of LR statistics indicate that the hypothesis "all slope parameters are zero" should be rejected at the level 0,05. In addition, in obtained models all explanatory variables are significant at 0,05.

Econometrical analysis has proved that the features which had the greatest influence on a risk being put into poverty, want or prosperity were:

- sort of the biological type of household,
- area of agricultural farm,

- farmer's level of education,
- standard of accommodation (number of rooms, hot running water, stationary telephone),
- province,
- number of sources of livelihood.

There have been many foreign researches pointing for the co-dependence of age of a head household and level of subjective poverty. Our research hasn't confirmed that.

After estimation the models we computed some predicted probabilities for households with particular set of characteristics. In table 3 we show such predictions for average[3] households from mazowieckie province with good living conditions (with hot water, telephone) and without jobless members due to illness.

Table 3. Predicted probabilities of exclusion with respect to education and gender – subjective approach

Level of education	In present income household meets two ends:			
	very poor	poor	insufficient	scarcely enough or good
Male head of household				
Lack of secondary	0,09	0,21	0,57	0,13
At least secondary	0,06	0,16	0,60	0,19
Female head of household				
Lack of secondary	0,13	0,27	0,51	0,09
At least secondary	0,09	0,21	0,57	0,13

Source: own calculations.

The analysis of the figures of the respective probabilities makes possible to compare the risk being put into poverty for different households. For instance, having in mind the household with an income hardly letting meet two ends there can be expected that it has been by 7% smaller to an average farmers' households where a head male farmer had been graduated from the secondary school, than in case of the farms dealt by women without such education.

Comparing results showed in table 2 we can see following main differences in obtained models:

- gender of head of household and number of jobless due to illness didn't affect significantly objective poverty, instead they did subjective poverty,
- the more were sources of livelihood the better was opinion of income situation up to the farmers, through the analysis of this situation proves something else,
- taking into account the statistical significance of explanatory variables, households in podlaskie and warmińsko-mazurskie provinces particularly were considered in subjective approach and households in podkarpackie and zachodniopomorskie – in subjective one.

Except to above differences conclusions derived from analyses of both models are similar.

[3] In that paper "average household" means hypothetical household with average level of quantitative attributes, moreover it isn't couple without children.

6. Conclusions

The carried out analysis under assumption *ceteris paribus* has made possible to come to the following conclusions.

- Increase in number of livelihood sources has brought about the better subjective evaluation of the farmers' household situation. The more numerous livelihood sources might develop the sense of bigger safety in farmers' opinion. The objective analysis of that situation led to quite different conclusions.
- The level of education affected the risk of getting to the extent of poverty explicitly both in subjective and objective aspects. That human potential proved to be essential in obtaining prosperity of farmers' households.
- Women as heads of households saw their income situation worse than men. Up to objective approach there wasn't any co-dependence of head gender and fact of getting to poverty or prosperity area.
- Married couples without children in farmers' households were in better situation than the other sorts of biological type of households.
- Higher accommodation standards referred usually to satisfied level of prosperity both in objective and subjective aspects.
- The farmers from Mazowsze province were less threatened by poverty than the ones from the other part of the country.
- Bigger number of dependants badly affected the fact of getting into prosperity area.
- Farmers households of bigger land area were less threatened than the smaller ones. The results may be essential having in mind the pending changes over the last years in the Polish country. There is a steady persisting polarization process leading to disappearing the medium-size farms for advantage of very big ones on one pole and very small on the other. It may result in increase of number of poor farms producing mainly for their own needs and rich big trade ones.

Finally, we can state that ordered models are useful tools for analytical, policy making and for monitoring purposes. They enable understanding the factors determining poverty and identification of vulnerable groups of households. Obtained results should be used in creation effective social policy.

References

Borooah V.K. (2001), *Logit and Probit. Ordered and Multinomial Models*, Sage Publications Inc., Thousand Oaks.

Greene W.H. (2000), *Econometric Analysis*, Prentice Hall Inc., Upper Saddle River.

Gruszczyński M. (2002), *Modele i prognozy zmiennych jakościowych w finansach i bankowości*, Oficyna Wydawnicza SGH, Warszawa.

Household Budget Surveys in 2003 (2004), Statistical Publishing Establishment, Warsaw.
Long J.S., Freese J. (2001), *Regression Models for Categorical Dependent Variables using Stata*, Stata Press Publication, College Station, Texas.
Panek T. (1991), *Demographic and Socio-Economic Determinants of Poverty: Case Study of Poland*, [in:] *Poverty Measurement for Economies in Eastern European Countries*, Polish Statistical Association, Warsaw.
Stata Base Reference Manual (2005), Vol. 2, Stata Press Publ., College Station, Texas.

IDENTYFIKACJA GOSPODARSTW DOMOWYCH ROLNIKÓW ZAGROŻONYCH UBÓSTWEM NA PODSTAWIE UPORZĄDKOWANEGO MODELU LOGITOWEGO

Streszczenie

W prezentowanej pracy podjęto temat finansowego aspektu dobrobytu w gospodarstwach domowych rolników. Uwzględniono zarówno subiektywne, jak i obiektywne ujęcie tego zagadnienia. Do identyfikacji gospodarstw o różnym stopniu zagrożenia ubóstwem i niedostatkiem wykorzystano wyniki otrzymane na podstawie uporządkowanych modeli logitowych. Uzyskane wyniki mogą mieć istotne znaczenie w aspekcie zmian dokonujących się w ostatnich latach na polskiej wsi.