המרכז למחקר בכלכלה חקלאית
THE CENTER FOR AGRICULTURAL ECONOMIC RESEARCH

Working Paper No. 9505

Estimation of an Endogenous Switching Regression Model with Discrete Dependent Variables: Monte-Carlo Analysis and Empirical Application of Three Estimators

by

Ayal Kimhi

Rehovot, Israel, P.O.B. 12

רחובות, ת.ד. 12

Working Paper No. 9505

Estimation of an Endogenous Switching Regression Model with Discrete Dependent Variables: Monte-Carlo Analysis and Empirical Application of Three Estimators

by

Ayal Kimhi

THE CENTER FOR AGRICULTURAL ECONOMIC RESEARCH
P.O. BOX 12, REHOVOT

Estimation of an Endogenous Switching Regression Model with Discrete Dependent Variables: Monte-Carlo Analysis and Empirical Application of Three Estimators*

Prepared for Presentation at the 7th World Congress of the Econometric Society, Tokyo, Japan, 22-29 August 1995

by

Ayal Kimhi

Department of Agricultural Economics and Management
Hebrew University, P.O. Box 12, Rehovot 76100, Israel

Summary

The performances of alternative two-stage estimators for the endogenous switching regression model with discrete dependent variables are compared, with regard to their usefulness as starting values for maximum likelihood estimation. This is especially important in the presence of large correlation coefficients, in which case maximum likelihood procedures have difficulties to converge. Monte-Carlo simulations indicate that an estimator that corrects for conditional heteroskedasticity of the residuals is superior in almost all instances, and especially when maximum likelihood is problematic. This result is also obtained in an empirical example in which off-farm work participation equations of farm women are conditional on farm work participation status.

* This research was supported by Grant No. IS-1845-90 from BARD, the United States - Israel Binational Agricultural Research and Development Fund, and initiated while I was visiting at the Department of Agricultural and Resource Economics, University of Maryland. I thank John Horowitz, Erik Lichtenberg and David Stapleton for helpful discussions, and Haim Regev from the Central Bureau of Statistics in Israel for his help with data issues. An earlier version of this paper appeared as AREC Working Paper No. 91-15, The University of Maryland. All errors are exclusively mine.

1 Introduction

This paper deals with estimating a special case of the endogenous switching regression model, in which all dependent variables are discrete. This case is a variation of the two-equation "Multivariate Probit Model with Structural Shift" described by Heckman (1978), in which structural shift exists in one equation only. Although writing down the likelihood function for this model is fairly straightforward, empirical applications might fail when using arbitrary starting values in the maximum likelihood estimation, especially when the correlation coefficients are large in absolute value. In other switching regression models this problem is solved by using a two-stage method to derive consistent starting values. This is not possible when all the dependent variables are discrete.

This paper considers two alternative two-stage procedures for the derivation of starting values to be used in maximum likelihood estimation. Neither procedure provides consistent estimators, but it is shown that at least one of them is preferred to using arbitrary starting values. Arbitrary starting values can be problematic in the presence of large (in absolute value) correlation coefficients. This statement is illustrated in figure 1, in which the likelihood function of the model used in this paper is simulated (for 10,000 observations). The two panels of the figure present two examples with different parameter vectors. It is easily observed that as the correlation coefficient increases in absolute value, it becomes harder to identify the true parameter.

The first procedure is the one suggested by Maddala (1983, pp. 223) for the traditional switching regression model: the first stage consists of estimating the selection equation, calculating selection correction terms and inserting them into the other equations. When the dependent variables are continuous, the second-stage equations are then consistently estimated by least squares. When the dependent variables are discrete, the second stage has to involve maximum likelihood methods. The resulting estimator is inconsistent since, by construction, the exact distribution of the stochastic terms is unknown. In addition, the stochastic terms are heteroskedastic (Yatchew and Griliches 1985), and hence this estimator will be denoted as the "heteroskedastic" estimator.

The second estimator uses the same first-stage estimators, and makes use of weighted observations in the second stage. Each observation is weighted by the inverse of its calculated standard deviation, which is a function of both first-stage and second-stage parameters. This corrects the heteroskedasticity problem, but the resulting probit model is still misspecified, rendering this estimator inconsistent as well. Nevertheless, it is a version of White's (1982) Quasi Maximum Likelihood estimator which possesses some desirable properties. It will be denoted as the "corrected" estimator.

The properties of the two alternative procedures are compared by a Monte-Carlo investigation. In particular, the estimators are compared to the maximum likelihood estimators under different correlation structures. Also, maximum likelihood estimation is performed using the two two-stage estimators as starting values, and their performance is compared in the two cases. Most of the findings indicate that the "corrected" estimator provides better starting values than the "heteroskedastic" estimator.

The suggested procedures are applied in estimating participation equations of farm women in farm work and off-farm work, when the off-farm participation equation parameters depend on farm participation. Again, the "corrected" estimator performs better than the "heteroskedastic" estimator, in the sense that it is closer to the maximum likelihood estimator.

The general model and the two-stage estimation procedures are described in section 2. Section 3 reports the results of the Monte-Carlo investigation. Section 4 develops the empirical application, and section 5 contains the results. Section 6 concludes.

2 THE MODEL AND THE TWO-STAGE ESTIMATION PROCEDURE

Heckman (1978) discusses the following model:

$$y_1^* = X_1 \cdot \alpha_1 + d_1 \cdot \beta_1 + y_2^* \cdot \gamma_1 + U_1 \qquad (1a)$$

$$y_2^* = X_2 \cdot \alpha_2 + d_1 \cdot \beta_2 + y_1^* \cdot \gamma_2 + U_2 \qquad (1b)$$

$$d_1 = 1 \quad \text{iff} \quad y_1^* > 0 \qquad (1c)$$

$$d_1 = 0 \quad \text{otherwise}$$

where U_i (i=1,2) is distributed as $N(0,\sigma_i^2)$, independently of X_i.

The endogenous switching regression model discussed by Maddala (1983, pp. 223) is derived from this model by assuming that y_1^* is unobserved, $\gamma_1=\gamma_2=\beta_1=0$, $\beta_2=X_3\cdot\alpha_3$, where $X_3 \supseteq X_2$, and that $U_2 = d_1\cdot U_{22} + (1-d_1)\cdot U_{21}$, where the assumptions regarding the U_i's apply to the U_{2i}'s as well. The resulting model, after some changes of notation ($X_{21}\cdot\alpha_{21} = X_2\cdot\alpha_2$; $X_{22}\cdot\alpha_{22} = X_2\cdot\alpha_2 + X_3\cdot\alpha_3$), is:

$$y_1^* = X_1\cdot\alpha_1 + U_1 \qquad (2a)$$

$$y_{21}^* = X_{21}\cdot\alpha_{21} + U_{21} \quad \text{iff} \quad d_1 = 0 \qquad (2b)$$

$$y_{22}^* = X_{22}\cdot\alpha_{22} + U_{22} \quad \text{iff} \quad d_1 = 1 \qquad (2c)$$

Assume further that the y_{2i}^*'s are unobserved. Instead, we observe d_{21} or d_{22} defined as:

$$\left.\begin{array}{lll} d_{21} = 1 & \text{iff} & y_{21}^* > 0 \\ d_{21} = 0 & \text{otherwise} & \end{array}\right\} \text{iff } d_1 = 0 \qquad (2d)$$

$$\left.\begin{array}{lll} d_{22} = 1 & \text{iff} & y_{22}^* > 0 \\ d_{22} = 0 & \text{otherwise} & \end{array}\right\} \text{iff } d_1 = 1$$

The log likelihood function of the model described in (2a)-(2d) is:

$$\ln L = \sum \{d_1\cdot d_{22}\cdot \ln\Phi(-A_1,-A_{22},\rho_2) + d_1\cdot(1-d_{22})\cdot\ln\Phi(-A_1,A_{22},-\rho_2) +$$
$$+(1-d_1)\cdot d_{21}\cdot\ln\Phi(A_1,-A_{21},\rho_1) + (1-d_1)\cdot(1-d_{21})\cdot\ln\Phi(A_1,A_{21},\rho_1)\}$$

where summation is over observations, ρ_i is the correlation coefficient between U_1 and U_{2i} (i=1,2), $A_1=-X_1\cdot\alpha_1/\sigma_1$, $A_{21}=-X_{21}\cdot\alpha_{21}/\sigma_{21}$, and $A_{22}=-X_{22}\cdot\alpha_{22}/\sigma_{22}$; σ_{21} and σ_{22} are the standard deviations of U_{21} and U_{22}, respectively, and Φ is the cumulative distribution function of a standardized bivariate normal random variable.

The first stage of Maddala's (1983) two-stage estimation method is unchanged by the assumption that the y_{2i}^*'s are unobserved: estimate (2a) by probit to get estimates of α_1/σ_1. The second stage, however, must also be estimated by probit, in each of the subsamples defined by $\{d_1=1\}$ and $\{d_1=0\}$, respectively. In order to correct for selectivity, we write, following Johnson and Kotz (1970, pp. 81), and assuming that $\sigma_1=\sigma_2=1$ (this is a convenient normalization since the standard errors are not identified):

$$U_{21} = \rho_1\cdot U_1 + u_1 \tag{3a}$$

$$U_{22} = \rho_2\cdot U_1 + u_2 \tag{3b}$$

$$E(U_1|d_1=0) = E(U_1|U_1<-X_1\cdot\alpha_1) = -\phi(-X_1\cdot\alpha_1)/\Phi(-X_1\cdot\alpha_1) \equiv \lambda_1 \tag{4a}$$

$$E(U_1|d_1=1) = E(U_1|U_1>-X_1\cdot\alpha_1) = \phi(-X_1\cdot\alpha_1)/(1-\Phi(-X_1\cdot\alpha_1)) \equiv \lambda_2 \tag{4b}$$

where the u_i's are independent of U_1, and ϕ is the density function of a standard normal random variable. Therefore:

$$E(U_{21}|d_1=0) = \rho_1\cdot\lambda_1 \tag{5a}$$

$$E(U_{22}|d_1=1) = \rho_2\cdot\lambda_2 \tag{5b}$$

Define:

$$\epsilon_1 = U_{21} - E(U_{21}|d_1=0) \tag{6a}$$

$$\epsilon_2 = U_{22} - E(U_{22}|d_1=1) \tag{6b}$$

and put into (2b) and (2c) to get:

$$y_{21}^* = X_{21} \cdot \alpha_{21} + \rho_1 \cdot \lambda_1 + \epsilon_1 \tag{7a}$$

$$y_{22}^* = X_{22} \cdot \alpha_{22} + \rho_2 \cdot \lambda_2 + \epsilon_2 \tag{7b}$$

where:

$$E(\epsilon_1|d_1=0) = 0 \tag{8a}$$

$$E(\epsilon_2|d_1=1) = 0 \tag{8b}$$

In Maddala's (1983) model, this is sufficient to get consistent estimators of α_{21} and α_{22}, after substituting the first-stage estimator $\hat{\alpha}_1$ for α_1. In our case, since probit is used in the second stage, one can only identify $\alpha_{2i}/[Var(\epsilon_i)]^{1/2}$ if $Var(\epsilon_i)$ is identical across observations. However, by construction, $Var(\epsilon_i)$ depends on X_1 (via λ_i), which varies across observations. ϵ_i is therefore heteroskedastic (Yatchew and Griliches 1985), and hence this estimator is denoted the "heteroskedastic estimator."

If $Var(\epsilon_i)$ could be calculated explicitly, one could correct the heteroskedasticity by using the normalized random variables $\epsilon_i/[Var(\epsilon_i)]^{1/2}$, which have a unit variance, instead of ϵ_i. Using Johnson and Kotz (1970, pp. 83), it can be shown that:

$$\text{Var}(U_1|d_1{=}0) = 1 + \lambda_1\cdot(-X_1\cdot\alpha_1 - \lambda_1) \tag{9a}$$

$$\text{Var}(U_1|d_1{=}1) = 1 + \lambda_2\cdot(-X_1\cdot\alpha_1 - \lambda_2) \tag{9b}$$

Using (3) and (6), and the fact that $Var(u_i) = 1 - \rho^2$:

$$\text{Var}(\epsilon_1|d_1{=}0) = \text{Var}(U_{21}|d_1{=}0) = 1 + \rho_1^2\cdot\lambda_1\cdot(-X_1\cdot\alpha_1 - \lambda_1) \equiv s_1^2 \tag{10a}$$

$$\text{Var}(\epsilon_2|d_1{=}1) = \text{Var}(U_{22}|d_1{=}1) = 1 + \rho_2^2\cdot\lambda_2\cdot(-X_1\cdot\alpha_1 - \lambda_2) \equiv s_2^2 \tag{10b}$$

Dividing (7a) and (7b) by s_1 and s_2, respectively, one gets the normalized second-stage equations, which are nonlinear in the parameters α_{2i} and ρ_i. Identification is supported by the following intuitive argument: conditional on ρ_i, α_{2i} is identified. It is then possible to estimate α_{2i}, given different values of ρ_i, and choose the one that results in the highest likelihood value. This depends, of course, on knowing the values of λ_1, λ_2, and α_1, and on the familiar condition that $X_{2i} \not\subset X_1$; $i=1,2$.

The problem is that λ_1, λ_2, and α_1 are unknown, and one has to substitute $\hat{\lambda}_1$, $\hat{\lambda}_2$, and $\hat{\alpha}_1$, respectively, where $\hat{\alpha}_1$ is the first-stage estimator. The resulting estimator is denoted as the "corrected" estimator. Since the distribution of the corrected residuals is unknown, the probit model serves as an approximation of the true probability model, and the probit estimator is a Quasi Maximum Likelihood Estimator (QMLE). White (1982) has shown that QMLE is a natural estimator of the parameter vector that minimizes the Kullback-Leibler Information Criterion. Therefore, it possesses some desirable properties. It could also be claimed that ϵ_i is not independent across observations, because of its dependence on $\hat{\lambda}_i$ which is calculated using all observations. However, if $\hat{\lambda}_i$ is consistent, ϵ_i can be said to be asymptotically independent, using an argument similar to that of Lee (1979).

The question is whether the "corrected" estimator is preferred over the "heteroskedastic" estimator as starting values for maximum likelihood estimation. It is not easy to derive the asymptotic biases of these estimators. Hence they will be compared by Monte-Carlo simulation methods. An empirical example will also be used to compare the two-stage estimators to the maximum likelihood estimator.

3 A MONTE-CARLO ANALYSIS

This section will present the results of various comparisons of the different estimators, derived using artificial data. The basic underlying model is (2), where each matrix of explanatory variables is composed of a unit vector and a uniformly distributed random vector defined over the interval [-0.5,0.5]. The following parameter vectors are used: $\alpha_{21} = \alpha_1 = (0,1)'$; $\alpha_{22} = (0,2)'$. The disturbance vectors are drawn from a standard normal distribution, where ρ_i is the correlation between U_1 and U_{2i}, i=1,2. The pair (ρ_1, ρ_2) take different values as described below.

Starting values for the six parameters are constructed as the true parameters plus a random error drawn from a $N(0,\sigma^2)$

distribution, where σ takes different values as described below. Starting values for the correlation coefficients are 0.001 in all cases. Each simulation is repeated 100 times, when the explanatory variables, disturbances, and starting values are redrawn each time. Six models are estimated in each repetition. Four of them use the arbitrary starting values: (1) the maximum likelihood model; (2) the first-stage probit model; (3) the "heteroskedastic" second-stage probit model; and (4) the "corrected second-stage probit model. The last two models, (5) and (6), are maximum likelihood models which use as starting values the "heteroskedastic" and the "corrected" estimators, respectively.

Figure 2 compares the parameter estimates of the slope coefficient in α_{21} for three different correlation structures and for four different sample sizes, when σ=0.1. Each estimate is in fact the average over the 100 repetitions, excluding those in which any of the models did not converge (see below). The maximum likelihood estimates are the averages of the three different maximum likelihood estimates derived in each repetition (in most cases these were practically the same). It can be seen that the parameter is overestimated in both the "heteroskedastic" procedure and the "corrected" procedure, but the deviation of the "corrected" estimate is much smaller. For both procedures the deviations increase with the absolute value of the correlation coefficients, and in most cases decrease with sample size. The sign of the correlation coefficients does not seem to matter. I have tried the combinations of {0.7,0.7}, {-0.7,-0.7}, {0.7,-0.7}, and {-0.7,0.7}, and the estimators did not behave differently in a noticeable way. In some cases, especially with the relatively small correlation coefficients, the "corrected" estimate was closer to the true parameter than the maximum likelihood estimate. Similar results are obtained when the estimates of the slope coefficient in α_{22} are compared.

Since the "corrected" estimates are closer to the maximum likelihood estimates than the "heteroskedastic" estimates, they should serve as better starting values for maximum likelihood estimation. To verify this claim, I compared in Table 1 the time that it took for maximum likelihood to converge, when the two sets of second-stage estimates are used as starting values alternatively. σ=0.1 was used here as well (a comparison of the number of iterations showed the same pattern as the comparison of time to convergence). It is again clear that the superiority of the "corrected" estimates is increasing with sample size. It is also increasing with the absolute value of the correlation coefficients for samples larger than 200 observations.

Another measure of the quality of starting values is the percent of cases in which estimation did not converge. In Table 1, the numbers in parentheses are the number of cases in which maximum likelihood converged using all the possible starting values. Convergence rates increase with sample size and decrease with the absolute value of the correlation coefficients. Figure 3 presents the number of cases in which estimation did not converge for each of the two starting values alternatives, in all the repetitions using 100 or 200 observations. The number of non-converging cases is smaller than the one seen in Table 1 since here, only cases in which any of the *slope* coefficients did not converge are considered. Although in both cases convergence decreases with the absolute value of the correlation coefficients, the superiority of the "corrected" starting values is evident, especially in the cases of relatively large (in absolute value) correlation coefficients.

It was mentioned earlier that the inconsistency of the "heteroskedastic" and the "corrected" second-stage estimators is a consequence of heteroskedasticity and deviations from normality, respectively, of the error terms. If this is true, it might be possible to decide which of the two estimators is preferred in a given setup by looking at the distributions of the relevant residuals and comparing them to the distribution of a standard normal random variable. The relevant residuals for the "heteroskedastic" estimation procedure are $\hat{\epsilon}_i$ which are derived from (7) by substituting $\hat{\lambda}_i$ for λ_i, $i=1,2$. In order to get the relevant residuals for the "corrected" estimation procedure, these are further divided by $\hat{s}_i$ which are derived similarly from (10). The true residuals are U_{2i} defined in (2). All three sets of residuals are calculated using simulations similar to those described above but with 20,000 observations each.

The actual and calculated residuals are compared in Figure 4 for two combinations of correlation coefficients, (0.7,0.7) and (0.9,0.9). For correlation coefficients that are smaller than 0.7 in absolute value, the three sets of residuals have very similar distributions. In both cases it is clear that the distribution of the "corrected" residuals is closer to the distribution of the true residuals than that of the "heteroskedastic" residuals. The differences between the three distributions increase with the absolute value of the correlation coefficients. This explains the fact that the "corrected" estimator provides better starting values than the "heteroskedastic" estimator. It also provides support to the claim that even though both second-stage estimators are inconsistent, the asymptotic bias of the "corrected" estimator is smaller than that of the "heteroskedastic" estimator.

Finally, the performance of the different starting values is compared when the standard deviation of starting values is changing. In particular, the simulations are repeated with 100 observations in each repetition, using correlation coefficients of (-0.7,0.7), with alternative values of σ (the standard deviation of the starting values). Maximum likelihood estimation is performed with three alternative sets of starting values: a random vector; the "corrected" estimates; and the "heteroskedastic" estimates (the same random vector of starting values is used in each of the two-stage procedures and in the maximum likelihood procedure).

The number of cases in which estimation did not converge in at least one of the three procedures is presented in Figure 5. As expected, the total number of cases is increasing with the standard deviation of starting values. This is also true when looking at the arbitrary starting values and the "heteroskedastic" estimates separately. On the other hand, the number of cases in which maximum likelihood did not converge when using the "corrected" estimates as starting values is decreasing with the standard deviation of starting values. It is also smaller than the number of cases in which each of the other two procedures did not converge, for all values of σ. This is another result in favor of using the "corrected" estimation procedure to generate starting values for maximum likelihood estimation.

4 AN EMPIRICAL EXAMPLE

To apply the econometric procedures described above, I use a model of farm women's off-farm work participation, in which the latent variable describing the tendency to participate depends on a farm participation dummy, which is also determined by a latent variable crossing a threshold.

The theoretical model assumes utility maximization over consumption and leisure subject to time and budget constraints, where time can be productively used on or off the farm (Kimhi 1994). Formally, the optimization problem is:

$$\max_{Th,C,Tf,Tm} U(Th, C) \qquad (11)$$

$$\text{s.t.} \quad \begin{aligned} &1. \quad C \le \pi(Tf) + W \cdot Tm + I \\ &2. \quad Th + Tf + Tm \le T \\ &3. \quad Tf \ge 0 \\ &4. \quad Tm \ge 0 \end{aligned}$$

where *Th*, *Tf* and *Tm* are time spent on home activities, farm work and off-farm work, respectively, *C* is consumption, *I* is unearned income, *W* is the off-farm wage and π is the conditional variable profit function described by Lopez (1982).

Two of the Kuhn-Tucker necessary conditions for maximization are:

$$\pi_1 + \delta/U_2 = U_1/U_2 \qquad (12a)$$

$$W + \phi/U_2 = U_1/U_2 \qquad (12b)$$

where δ and ϕ are positive if and only if farm work and off-farm work, respectively, are zero, and subscripts denote partial derivatives. If one proceeds to solve (12a) and (12b) simultaneously, he ends up with a common simultaneous equations model. In order to get the switching regression structure, I assume that for some reason (long-run considerations, etc.) the farm participation problem is solved prior to the off-farm participation problem. This results in the model described by (2), and therefore farm participation is determined solely by (12a). This ordering of decisions is absolutely arbitrary, since there are arguments for and against each of the two possible orderings. The ordering of decisions does not contradict the possibility that the stochastic terms U_{2i} (i=1,2) are drawn before the farm participation decision is made, which is necessary for the validity of the switching regression model (Poirier and Ruud 1981).

For those who work on the farm, $\delta=0$, $\pi_1 = U_1/U_2$, and off-farm participation occurs if:

$$W > \pi_1(Tf^*), \tag{13a}$$

assuming all sufficient conditions are met, where Tf^* denotes optimal farm labor supply given no off-farm work. For those who do not work on the farm, off-farm participation occurs if:

$$W > U_1(I,T)/U_2(I,T). \tag{13b}$$

This leads to the following off-farm participation index function:

$$Y^* = W - RHS \tag{14}$$

$$Y = \begin{cases} 1 & \text{if } Y^* > 0 \\ 0 & \text{otherwise} \end{cases}$$

where *RHS* is the right-hand side of equations (13a) and (13b) for farm participants and nonparticipants, respectively. It is clear that when specifying Y^* as a function of observable variables, this function depends on farm participation.

In terms of the general model (2), this model is expressed as:

$$Y_1^* = \pi_1(0) - U_1(I,T)/U_2(I,T) \tag{2a}'$$

$$Y_{21}^* = W - U_1(I,T)/U_2(I,T) \tag{2b}'$$

$$Y_{22}^* = W - \pi_1(Tf^*), \tag{2c}'$$

and (2d). I specify these unobserved latent variables as linear combinations of explanatory variables, including personal, family and farm characteristics, and use data on farm women from the 1981 Census of Agriculture in Israel to estimate the model. Descriptive statistics of the data set are reported in Table 2. I chose farm women since they were more equally divided by farm participation status: only 10% of farm men did not work on the farm (Kimhi 1991), as compared to 59% of farm women (Table 2).

5 RESULTS

The results of the two-stage estimation procedures are compared to maximum likelihood results in Tables 3 and 4. In practice, maximum likelihood estimation was not much more time-consuming than the "corrected" probit two-stage procedure (the "heteroskedastic" procedure was faster than the other two). However, because the correlation coefficient was close to minus one in one of the subsamples, maximum likelihood estimation failed when arbitrary starting values were used. Maximum likelihood estimation was successful only when the two-stage estimators were used as starting values, demonstrating the usefulness of the two-stage procedure in this particular case.

Table 3 presents the results of the farm participation equation. The first column presents univariate probit estimates, the second, the maximum likelihood estimates (joint estimation with off-farm participation). Besides the variables shown in the table, the farm participation equation also included a set of regional dummies and a set of village establishment year dummies. The former set should have actually been included in the off-farm participation equation as well (Tokle and Huffman 1991), but it was excluded in this case for the purpose of identification of the second-stage parameters, as discussed in section 1.

One can see that the difference between the estimators is only marginal, which is not surprising, since univariate probit is a consistent QMLE in this case (Avery, Hansen and Hotz 1983). Age profiles of farm participation are concave as expected, with participation probability peaking around the age of 47. Schooling has a positive and significant effect. The number of other family members in all age groups affects farm participation negatively, with adults having a greater effect than children. In dairy farms, farm participation of farm women is much higher, which is expected, since dairy farm work is known to be a good complement to housework. Land size has a negative effect on participation: in larger farms, women have a lower tendency to work on the farm. It could be that hired labor substitutes for family members in larger farms, and the income effect may play a role here too. In contrast, capital stock has a positive effect on farm participation.

The capital stock variable includes only capital assets which were at least 10 years old, to avoid the problem of endogeneity of capital stock in the time allocation decision. Also, the land variable is the original land allotment of the farm, which was institutionally determined at the time of establishment of the village. The dairy farm dummy is also considered exogenous since strict milk quotas and large subsidies have kept the subset of farms that produce milk fairly stable over time (because of the endogeneity problem, the number of milk cows and other farm attributes were not used as explanatory variables).

Table 4 includes the off-farm participation results. Comparing the "heteroskedastic" and the "corrected" two-stage estimators, one can see that they are almost identical in the subsample of those who worked on the farm. However, in the subsample of those who did not, there is a difference between the two. Comparing them to the maximum likelihood estimators, it is evident that the "corrected" probit estimator is closer to the maximum likelihood estimator than the "heteroskedastic" one (with the exception of three coefficients only).

Comparing the off-farm participation equations for the two subsamples of farm women, we first notice that the correlation coefficient between the stochastic terms of the farm and off-farm participation equations is close to minus one in the equations of farm nonworkers, in contrast to a negative but insignificant correlation coefficient for farm workers. Elsewhere, the reverse result has been obtained for farm men (Kimhi 1991).

The coefficients of personal characteristics in the two subsamples are not significantly different. Off-farm participation probability as a function of age peaks slightly later for farm workers (at the age of 35 versus 33 for nonworkers), and in both cases off-farm participation probability peaks much earlier than farm participation probability (at the age of 47), and declines much faster afterwards. The schooling coefficient is positive and significant, and is approximately twice as large as the schooling coefficient in the farm participation equation, which means that schooling, at least as measured here, contributes more to off-farm earnings than to farm productivity. These results are very much in line with those of others (Lass, Findeis and Hallberg 1991), except that age profiles of off-farm participation probability are more concave and peak earlier than in the other studies (between ages 45 and 55).

The number of children decreases off-farm participation probability, and the number of adults increases it, in both subsamples. These effects are stronger in the nonworkers equation (with the exception of family members over 65 years of age). The major difference between the two subsamples lies in the coefficients of farm attributes. The land size variable has a negative coefficient for farm workers and a positive one for nonworkers. The dairy farm dummy has a negative coefficient in both cases, which is much larger in absolute value for farm workers. Capital stock has a negative and significant coefficient in the nonworkers equation, and a positive and non-significant coefficient in the workers equation. These differences are expected (Kimhi 1991), since for those who work on the farm, farm attributes affect farm labor demand and hence affect off-farm labor supply through the time constraint. For those who do not work on the farm, the effect is only through the budget constraint (i.e., both substitution and income effects exist for farm workers, while only the income effect exists for nonworkers). It is evident that the substitution and income effects work in the same direction in the workers' equation, since in dairy farms (where family labor demand is relatively higher), farm women have a higher tendency to work on the farm. This last finding is in line with farm men's participation results reported elsewhere (Lass et al. 1991). I do not have an explanation for the positive land coefficient in the nonworkers equation.

Finally, an attempt was made to evaluate the unconditional marginal effects of the explanatory variables on the latent off-farm participation patterns of farm women (Huang, Raunikar and Misra 1991; Kimhi 1992), rather than the partial effect represented by the estimated coefficients. The results are qualitatively unchanged with respect to personal and family characteristics. Land size has a small positive effect on the marginal off-farm participation tendency, in this case.

6 SUMMARY AND CONCLUSIONS

Good starting values for maximum likelihood estimation of an endogenous switching regression model with discrete dependent variables are particularly important when the correlations between the equation residuals are large in absolute value. This was demonstrated in the empirical application reported in this paper, in which maximum likelihood estimation could not converge when arbitrary starting values were used.

The traditional two-step procedure for estimating an endogenous switching regression model is unsuitable when only qualitative realizations of the dependent variables are observed. This is because after correcting for selectivity, the conditional distribution of the residuals is unknown. In the alternative method proposed in this paper, the residual is normalized by its calculated (observation-specific) standard deviation, which is a function of both first- and second-stage parameters. In this case the second-stage regressions are nonlinear in the parameters.

Although this "corrected" second-stage estimator is inconsistent as well for the same reason, it performed better than the 'naive' heteroskedastic estimator in a series of simulations performed with various sample sizes, correlation coefficients, and starting values. The distribution of the "corrected" residuals was found closer to that of a standard normal random variable than the distribution of the "heteroskedastic" residuals. As a result, the "corrected" estimates were closer to the maximum likelihood estimates, maximum likelihood converged faster when the "corrected" estimates were used as starting values, and fewer cases of nonconvergence were observed.

In the empirical example, in which a model of farm and off-farm work participation of farm women was estimated, the "corrected" estimates were quite close to the maximum likelihood estimates. This provides additional support for using the proposed "corrected" two-stage estimator for providing starting values for maximum likelihood estimation of such models. Further investigation of the performance of this estimator should include calculations of its asymptotic bias and attempts to correct the bias and/or characterize the conditions under which it is significant.

REFERENCES

Avery, R. B., Hansen, L. P., and Hotz, V. J. (1983), "Multiperiod Probit Models and Orthogonality Condition Estimation," International Economic Review, 24, 21-35.

Heckman, J. J. (1978), "Dummy Endogenous Variables in a Simultaneous Equation System." Econometrica, 46, 931-959.

Huang, C. L., Raunikar, R., and Misra, S. (1991), "The Application and Economic Interpretation of Selectivity Models," American Journal of Agricultural Economics, 73, 496-501.

Johnson, N. L., and Kotz, S. (1970), Distributions in Statistics: Continuous Univariate Distributions-1, Boston: Houghton Mifflin.

Kimhi, A. (1994), "Quasi Maximum Likelihood Estimation of Multivariate Probit Models: Estimating Farm Couples' Labor Participation," American Journal of Agricultural Economics, 76, 828-835.

_____ (1992), "The Application and Economic Interpretation of Selectivity Models: Comment," American Journal of Agricultural Economics, 74, 498-499.

_____ (1991), The Relevance of the Extent of Farm Work to the Analysis of Off-Farm Labor Supply of Farmers, Working Paper No. 91-11, Department of Agricultural and Resource Economics, University of Maryland.

Lass, D. A., Findeis, J. L., and Hallberg, M.C. (1991), "Factors Affecting the Supply of Off-Farm Labor: A Review of Empirical Evidence," in Multiple Job-Holding Among Farm Families, eds. M. C. Hallberg, J. L. Findeis, and D. A. Lass, Ames: Iowa State University Press, pp. 239-622.

Lee, L. F. (1979), "Identification and Estimation in Binary Choice Models with Limited (Censored) Dependent Variables," Econometrica, 47, 977-995.

Lopez, R. E. (1982), "Applications of Duality Theory to Agriculture," Western Journal of Agricultural Economics, 7, 353-365.

Maddala, G. S. (1983), Limited-Dependent and Qualitative Variables in Econometrics, Cambridge: Cambridge University Press.

Murphy, K. M., and Topel, R. H. (1985), "Estimation and Inference in Two-Step Econometric Models," Journal of Business and Economic Statistics, 3, 370-379.

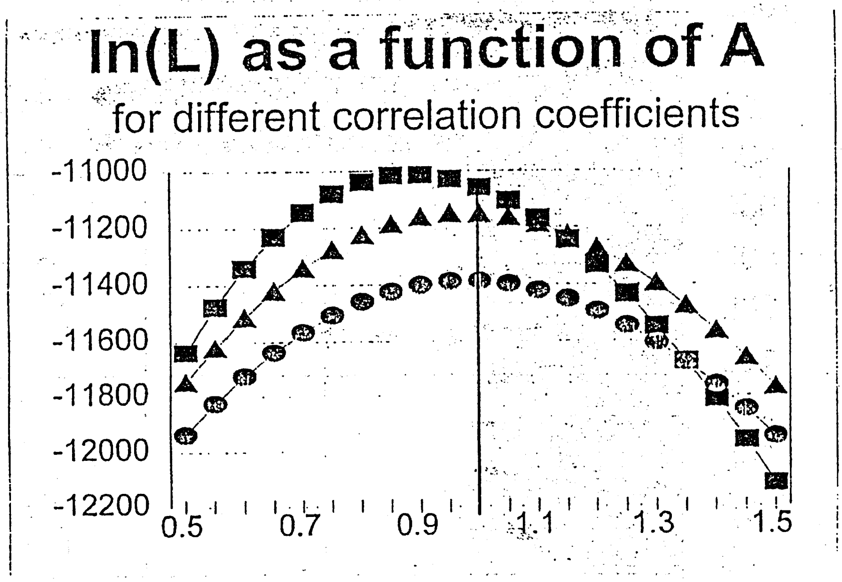
Poirier, D. J., and Ruud, P. A. (1981), "On the Appropriateness of Endogenous Switching," Journal of Econometrics, 16, 249-256.

Tokle, J. G., and Huffman, W. E. (1991), Local Economic Conditions and Wage Labor Decisions of Farm and Rural Nonfarm Couples," American Journal of Agricultural Economics, 73, 652-670.

White, H. (1982), "Maximum Likelihood Estimation of Misspecified Models," Econometrica, 50, 1-25.

Yatchew, A., and Griliches, Z. (1985), "Specification Error in Probit Models," The Review of Economics and Statistics, 67, 134-139.

a. Identification of A=1 when B=0



b. Identification of A=-2 when B=-1

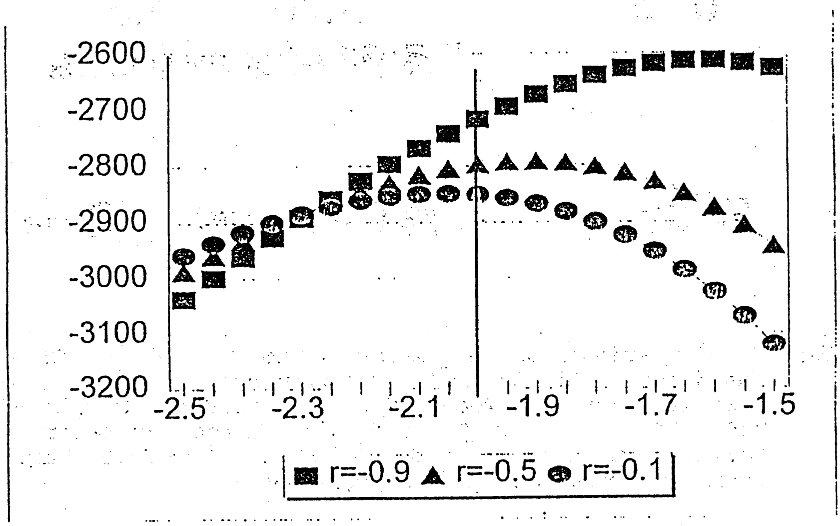


Fig.1. Identification in a Bivariate Probit model with parameters (A,B,r)

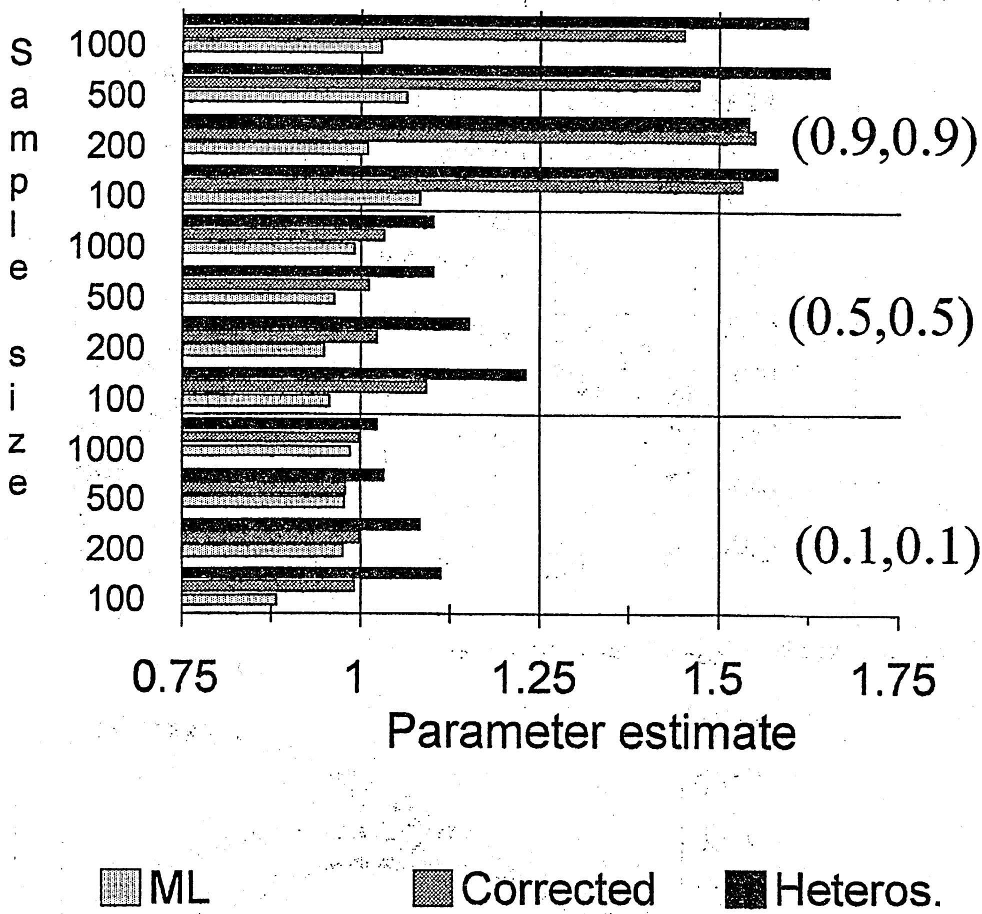


Fig. 2. Comparison of Parameter Estimates of the Three Methods (Maximum Likelihood; "Corrected" Two-Stage; and "Heteroskedastic" Two-Stage) for Different Sample Sizes (100, 200, 500, and 1000) and for Three Different Pairs of Correlation Coefficients ({0.1,0.1}, {0.5,0.5}, and {0.9,0.9} when the True Coefficient Equals One

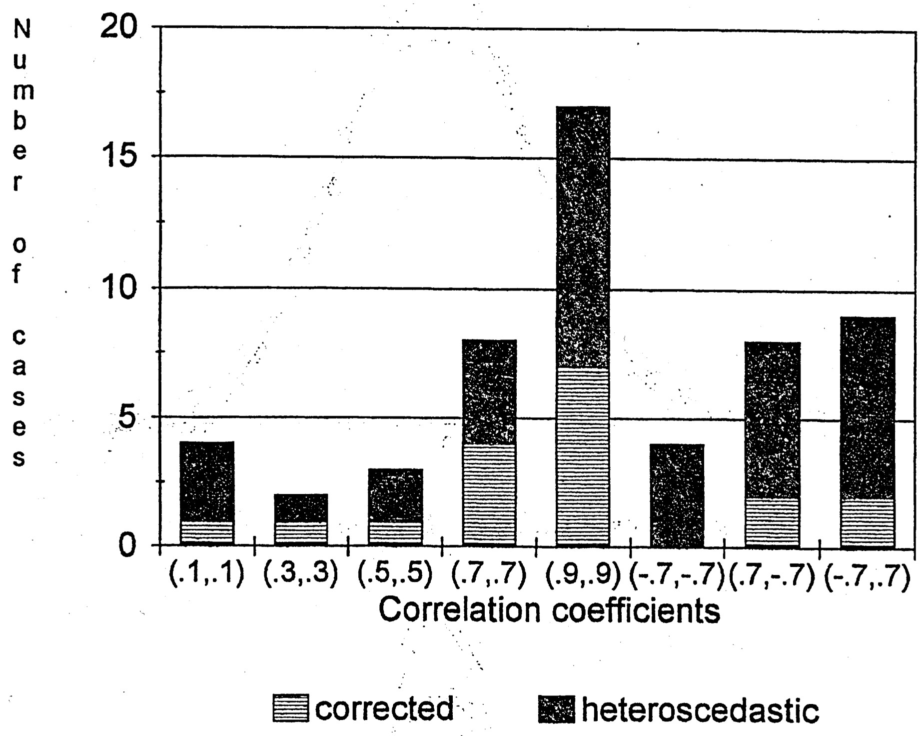
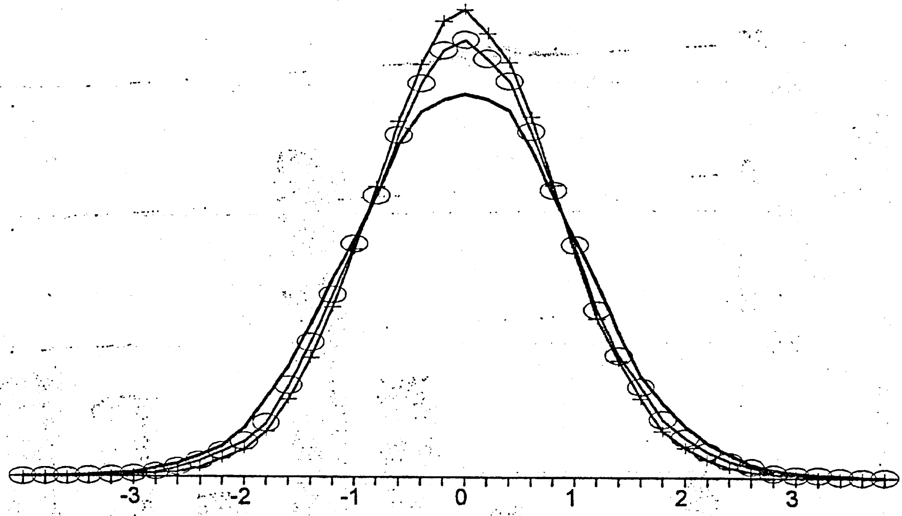
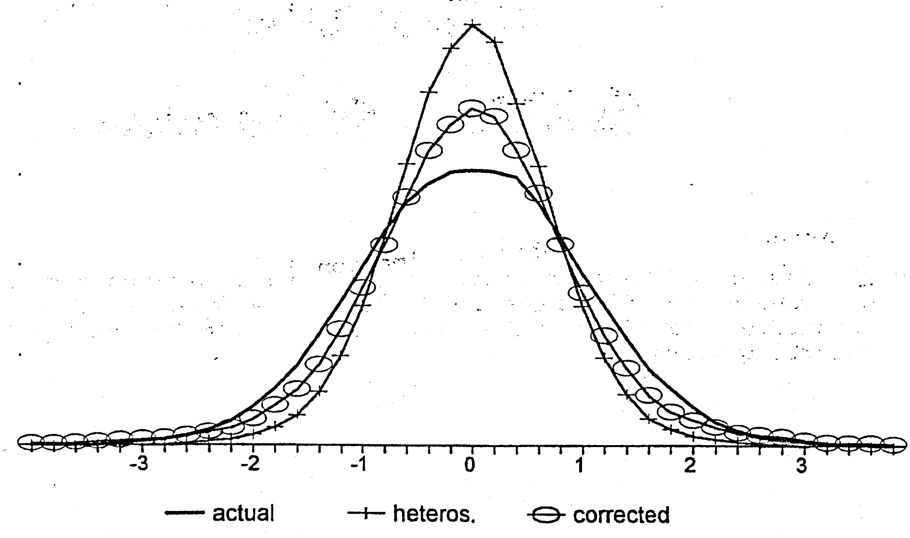


Fig. 3. Number of Cases in Which Estimation did not Converge, for Various Pairs of Correlation Coefficients, When the Cases are Separated to Those with "Corrected" or "Heteroskedastic" Starting Values.



a. Correlation coefficients (0.7,0.7)



b. Correlation coefficients (0.9,0.9)

Fig. 4. Distribution of Calculated Versus Actual Residuals Using Two Alternative Pairs of Correlation Coefficients

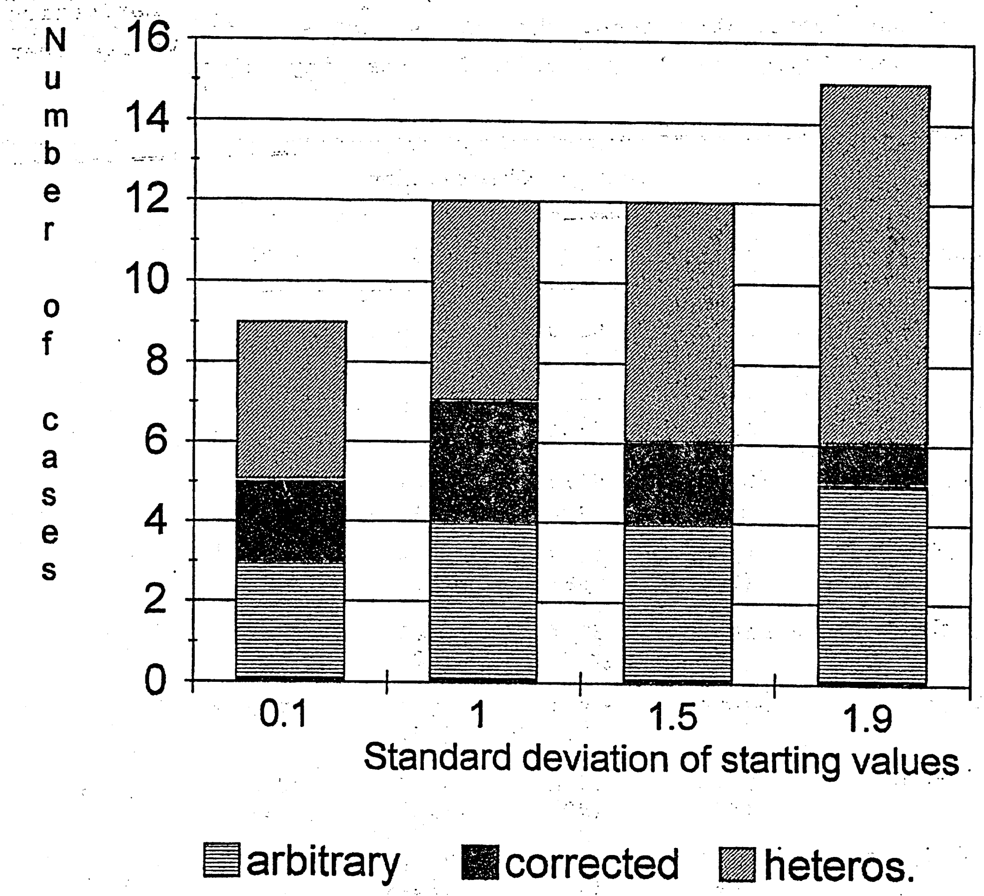


Fig. 5. Number of Cases in Which Maximum Likelihood Estimation did not Converge by Standard Deviation of Starting Values, When the Cases are Separated to Those Using Arbitrary, "Corrected", and "Heteroskedastic" starting values

Table 1. Percent of Cases in Which Maximum Likelihood Estimation was Faster when Using the "Corrected" Starting Values than when Using the "Heteroscedastic" Starting Values

	Number of Observations			
Correlations	100	200	500	1000
0.1,0.1	57.0 (84)	62.0 (95)	51.0 (99)	38.5 (96)
0.3,0.3	46.3 (82)	55.6 (90)	67.0 (91)	56.8 (95)
0.5,0.5	56.6 (76)	69.3 (88)	68.1 (91)	73.9 (88)
0.7,0.7	61.2 (85)	63.6 (88)	78.9 (90)	85.9 (92)
0.9,0.9	40.3 (77)	52.4 (82)	69.5 (82)	82.6 (86)
-0.7,-0.7	55.6 (81)	59.1 (93)	73.7 (95)	87.2 (94)
0.7,-0.7	58.8 (80)	54.4 (90)	84.6 (91)	78.3 (92)
-0.7,0.7	47.6 (84)	57.5 (87)	73.4 (94)	84.8 (92)

NOTE: Number of cases compared in parentheses. Excluded are all the cases in which at least one of the maximum likelihood estimations did not converge in 200 iterations.

Table 2. Descriptive Statistics

I. Explanatory Variables

Variable	*Mean (by Farm Participation)* *All*	*Workers*	*Nonworkers*	*Range*	*Units*
Age	43.6	43.0	44.0	14-80	years
Schooling	8.6	9.4	8.1	0-20	years
Family 0-14[a]	1.6	1.6	1.7	0-11	persons
Family 15-21	.88	.78	.95	0-8	persons
Family 22-65	2.4	2.2	2.4	0-10	persons
Family 66+	.23	.19	.26	0-3	persons
Total Land[b]	3.1	3.0	3.2	0-8	ln(dunams)[c]
Dairy Farm	.08	.09	.07	0-1	dummy
Old Capital[d]	1.0	1.1	.97	0-7.3	ln($81)[e]

II. Participation

	Working Off-Farm	*Not Working*	*Total*
Working on Farm	1075 (6%)	5866 (35%)	6941 (41%)
Not Working	2580 (15%)	7494 (44%)	10074 (59%)
Total	3655 (21%)	13360 (79%)	17015 (100%)

[a] Number of family members in each age group.
[b] Original land allotment.
[c] 1 dunam = 0.23 acre.
[d] Normative value of capital assets at least 10 years of age.
[e] In 1981 prices.

Table 3. Two-Stage and Maximum Likelihood Estimators of Farm Participation

Variable	*Two-Stage*	*Maximum Likelihood*
Intercept	-2.07 (-13.0)	-2.07 (-14.0)
Age	.086 (13.0)	.085 (13.0)
Age squared	-.0009 (-12.0)	-.0009 (-13.0)
Schooling	.034 (12.0)	.036 (13.0)
Family 0-14	-.025 (-3.3)	-.022 (-2.9)
Family 15-21	-.045 (-4.5)	-.045 (-4.5)
Family 22-65	-.105 (-9.7)	-.104 (-9.8)
Family 65+	-.099 (-3.9)	-.095 (-3.9)
Land[a]	-.036 (-2.7)	-.039 (-3.1)
Dairy Farm	.354 (9.2)	.340 (8.9)
Old Capital[a]	.025 (4.5)	.024 (4.2)

NOTE: asymptotic T-statistics in parentheses. Standard errors of the two-stage estimators were calculated using the method of Murphy and Topel (1985). The farm equation also included sets of regional dummies and village establishment year dummies.

[a] Land and capital stock were measured in natural logarithms to minimize the effects of outliers. Normalization was used such that a zero remained a zero.

Table 4. A Comparison of Estimators of the Off-Farm Participation Equations

	Farm Workers			*Nonworkers*		
	Two-Stage			*Two-Stage*		
Variable	*Heteros.*	*Corrected*	*ML*	*Heteros.*	*Corrected*	*ML*
Intercept	-2.73 (-7.0)	-2.72 (-6.9)	-2.52 (-6.1)	-3.82 (-16.0)	-2.98 (-13.0)	-3.14 (-17.0)
Age	.091 (5.5)	.091 (5.5)	.085 (5.0)	.099 (7.5)	.073 (6.9)	.078 (8.6)
Age squared	-.0013 (-6.6)	-.0013 (-6.7)	-.0012 (-6.2)	-.0015 (-9.5)	-.0011 (-8.7)	-.0012 (-11.0)
Schooling	.083 (10.0)	.082 (13.0)	.078 (11.0)	.109 (19.0)	.082 (14.0)	.087 (23.0)
Family 0-14	-.067 (-4.1)	-.066 (-4.1)	-.065 (-4.1)	-.113 (-9.3)	-.087 (-8.4)	-.090 (-9.0)
Family 15-21	.012 (.54)	.012 (.54)	.014 (.66)	-.016 (-.99)	-.014 (-1.1)	-.019 (-1.3)
Family 22-65	.022 (.87)	.021 (.90)	.024 (1.0)	.062 (3.2)	.045 (3.1)	.043 (2.9)
Family 65+	.182 (3.7)	.179 (3.6)	.181 (3.6)	.205 (4.9)	.154 (4.2)	.157 (4.7)
Land	-.108 (-4.6)	-.108 (-4.2)	-.098 (-3.8)	.094 (4.7)	.079 (5.2)	.091 (6.2)
Dairy Farm	-.444 (-4.8)	-.438 (-4.9)	-.449 (-5.1)	-.196 (-2.6)	-.160 (-3.0)	-.177 (-3.3)
Old Capital	.006 (.53)	.006 (.52)	.005 (.42)	-.044 (-4.8)	-.034 (-4.5)	-.035 (-4.9)
Correlation	-.202 (-1.6)	-.187 (-1.6)	-.270 (-2.1)	-.907 (-7.9)	-.860 (-17.0)	-.925 (-3.7)

NOTE: asymptotic T-statistics in parentheses. Standard errors of the two-stage estimators were calculated using the method of Murphy and Topel (1985).

PREVIOUS WORKING PAPERS

8701 Dan Yaron, Ariel Dinar and Aharon Ratner - **The Applicability and Usefulness of Cooperative Game Theory in the Analysis of Equity Issues in Regional Water Resource Problems.**

8702 Yoav Kislev and Arie Marvid - **Mishorim--A Regional Cooperative of Moshavim** (Hebrew).

8703 Eli Feinerman and Paul B. Siegel - **A Dynamic Farm Level Planning Model for Optimal Feedlot Production and Marketing: An Illustration for a Situation in Israel.**

8704 David Bigman - **Optimal Provision of Public Goods: Normative and Bargaining Solutions.**

8705 David Bigman and Paul D. McNelis - **Indexation, Contract Length and Wage Dispersion under Rapid Inflation: The Israeli Experience 1979-1984.**

8706 Haim Levy and Zvi Lerman - **Testing the Predictive Power of Ex-Post Efficient Portfolios.**

8707 Haim Levy and Zvi Lerman - **Internationally Diversified Bond and Stock Portfolios.**

8801 David Bigman, Nir Becker and Hector Barak - **An Economic Analysis of Wheat Growing in the Northern Negev and an Evaluation of the Drought Compensation Program** (Hebrew).

8802 Csaba Csaki - **Hungarian Agricultural Policy in the 80's: Economic Management and Organization of the Hungarian Agriculture.**

8901 Arye Volk - **Factors Affecting Growth of Debt on the Family Farms in a Moshav** (Hebrew).

8902 Yoav Kislev, Zvi Lerman & Pinhas Zusman - **Experience with Credit Cooperatives in Israeli Agriculture.**

8903 Zvi Lerman - **Capital Structure of Agricultural Cooperatives in Israel.**

8904 Yoav Kislev, Zvi Lerman and Pinhas Zusman - **Credit in the Moshav--Experience and Lessons** (Hebrew).

8905 Pinhas Zusman - **A Conceptual Framework for the Economy of the Moshav and its Structure** (Hebrew).

8906 Shlomo Yitzhaki - **On the Sensitivity of a Regression Coefficient to Monotonic Transformations.**

8907 Yair Mundlak - **Sources of Input Variations and the (In)Efficiency of Empirical Dual Production Functions.**

8908 Eli Feinerman and Meira S. Falkovitz - **An Agricultural Multipurpose Service Cooperative: Pareto Optimality, Price-Tax Solution and Stability.**

9001 Zvi Lerman and Claudia Parliament - **Performance of U.S. Agricultural Cooperatives: Size and Industry Effects.**

9002 Yoav Kislev - **The Economic Organization of Citrus Production in Israel** (Hebrew).

9003 Zvi Lerman and Claudia Parliament - **Comparative Performance of Food-Processing Cooperatives and Investor-Owned Firms in the U.S.A.**

9004 Alan Swinbank - **Europe After 1992 and Implications for Fresh Produce From Israel.**

9005 Ziv Bar-Shira - **A Non-Parametric Test of the Expected Utility Hypothesis.**

9006 Yoav Kislev - **The Water Economy of Israel** (Hebrew).

9101 Yoav Kislev and Willis Peterson - **Economies of Scale in Agriculture: A Reexamination of the Evidence.**

9102 van Dijk G. and C.P. Veerman - **The Philosophy and Practice of Dutch Co-operative Marketing.**

9103 Eli Feinerman and Ariel Dinar - **Economic and Managerial Aspects of Irrigation with Saline Water: The Israeli Experience.**

9104 Yoav Kislev - **Family Farms, Cooperatives, and Collectives.**

9105 Pinhas Zusman and Gordon C. Rausser - **Organizational Equilibrium and the Optimality of Collective Action.**

9106 Yoav Kislev - **The Economics of Water Resources - Principles and their Application** (Hebrew).

9107 Dan Yaron, Ariel Dinar and Hillary Voet - **Innovations on Family Farms: The Case of the Nazareth Region in Israel.**

9108 Pinhas Zusman - **A Conceptual Framework for a Regulatory Policy of the Israeli Water Resources** (Hebrew).

9109 Eitan Hochman and Oded Hochman - **A Policy of Efficient Water Pricing in Israel.**

9110 Dan Yaron - **Water Quota Allocation and Pricing Policy in Agriculture** (Hebrew).

9201 Yujiro Hayami - **Conditions of Agricultural Diversification for Economic Development.**

9202 Pinhas Zusman and Gordon C. Rausser -**Endogenous Policy Theory: The Political Structure and Policy Formation.**

9203 Domingo Cavallo - **Argentina's Recent Economic Reform in the Light of Mundlak's Sectorial Growth Model.**

9204 Pinhas Zusman - **Participants' Ethical Attitudes and Organizational Structure and Performance.**

9205 Pinhas Zusman - **Membership Ethical Attitudes and the Performance and Survivability of the Cooperative Enterprise.**

9206 Yoav Kislev - **The Cooperative Experience in Agriculture: International Comparisons.**

9207 Robert M. Behr - **Development and Prospects of World Citrus Markets.**

9208 Zvi Lerman and Claudia Parliament - **Financing of Growth in Agricultural Cooperatives.**

9209 Claudia Parliament and Zvi Lerman - **Risk and Equity in Agricultural Cooperatives.**

9210 Csaba Csaki and Zvi Lerman - **Land Reform and Farm Sector Restructuring in the Former Soviet Union and Russia.**

9301 Zvi Lerman, Evgenii Tankhilevich, Kirill Mozhin and Natalya Sapova - **Self-Sustainability of Subsidiary Household Farms: Lessons for Privatization in Russian Agriculture.**

9302 Ayal Kimhi - **Optimal Timing of Transferring the Family Farm from Father to Son.**

9303 Ayal Kimhi - **Investment in Children, Selective Off-Farm Migration, and Human Capital of Future Farmers: A Dynastic Utility Model.**

9304 Meira S. Falkovitz and Eli Feinerman - **Optimal Scheduling of Nitrogen Fertilization and Irrigation.**

9305 Shlomit Karidi-Arbel and Yoav Kislev - **Subsidies and Planning in Broilers--the 1980s.** (Hebrew).

9401 Karen Brooks and Zvi Lerman - **Restructuring of Socialized Farms and New Land Relations in Russia.**

9402 Karen Brooks and Zvi Lerman - **Changing Land Relations and Farming Structures in Former Socialist Countries.**

9403 Vardit Heber and Yoav Kislev - **The Protection of Domestic Production in Agriculture - Three Products.** (Hebrew).

9404 Eyal Brill and Eithan Hochman - **Allocation and Pricing of Water Under Common Property at the Regional Level** (Hebrew).

9405 Yacov Tsur and Eithan Hochman - **The Time to Adopt: Count-Data Regression Models of Technology Adoption Decisions.**

9406 Shuky Regev - **Farm Succession--The Legal Aspect.** (Hebrew).

9407 Yoav Kislev - **A Statistical Atlas of Agriculture in Israel, 1994 Edition** (Hebrew).

9408 Eithan Hochman and Eyal Brill - **Homogeneity and Heterogeneity of Bio-Resources in Economic Models.**

9409 Eyal Brill and Eithan Hochman - **Allocation and Pricing of Water with Common Property at the Regional Level.**

9501 Kwang ho Cho - **An Economic Diagnosis and Decision Model of Dairy Farming.**

9502 Shuky Regev - **Farm Succesion in the Moshav - A Legal Examination.**

9503 Naomi Nevo - **Inter-Generational Transfer of Farms in Moshavei-Ovdim: An Ethnographical Enquiry.**

9504 Menahem Kantor - **Water Issues in Israel Towards the Next Century.** (Hebrew).

9505 Ayal Kimhi - **Estimation of an Endogenous Switching Regression Model with Discrete Dependent Variables: Monte-Carlo Analysis and Empirical application of Three Estimators.**