ESTIMATION OF SOME LIMITED DEPENDENT VARIABLE MODELS WITH APPLICATION TO HOUSING DEMAND

by

Lung-Fei Lee and R. P. Trost*

Discussion Paper No. 77-84, June 1977

---

*The authors are affiliated with the University of Minnesota and University of Florida, respectively. We would like to thank G. S. Maddala, Jerome W. Milliman, James Heckman, F. D. Nelson and R. Blane Roberts for many helpful comments. Responsibility for any error is solely ours. Financial support from the National Science Foundation under grant SOC-76-04356 to the University of Florida is gratefully acknowledged.

Center for Economic Research
Department of Economics
University of Minnesota
Minneapolis, Minnesota 55455

I. Introduction

The first limited dependent variable model was proposed by Tobin [23]. His approach is now known as Tobit analysis. Tobit analysis uses a maximum likelihood procedure to estimate model with a truncated dependent variable when a large number of observations take on the truncating threshold. Tobin's model has recently been extended to handle more complicated situations. Heckman [10] and Nelson [16] employ censored dependent variable models to estimate labor supply curves. Other extentions of Tobin's paper include econometric studies of disequilibrium market models. For examples of these see Fair and Jaffee [5], Maddala and Nelson [14] and Goldfeld and Quandt [8], among others. In general, the structures of these models are highly nonlinear and maximum likelihood estimation methods are suggested. But the successful utilization of these maximum likelihood procedures is crucially dependent upon available numerical iterative algorithms. While numerical algorithms such as the Newton-Raphson iteration procedure are quite successful for the simple Tobit model, for more complicated models straightforward utilization of numerical iterative algorithms can be quite expensive. For example, Nelson [16] reports that the Newton-Raphson and Quadratic hill climbing algorithms cannot be easily applied to the Censored Regression Models without good initial estimates.

One purpose of this paper is to suggest a simple method of obtaining these good initial estimates for a certain class of limited dependent variable models. More specifically, we propose some initial consistent estimates for a model that extends the switching regression model of Goldfeld and Quandt [7]. We point out that our model combines several different limited dependent variable models into one general framework. By starting with these initial consistent estimates, a two step maximum likelihood procedure can be applied which simplifies the attainment of maximum likelihood estimates.

A second purpose of this paper is demonstrating how our model and estimation techniques can be used to study housing demand. We argue that previous studies on this subject only emphasize one part of the complete model. Some of these previous studies only consider demand (or expenditure) equations for owners and renters, while others only consider the consumer's rent versus own decision. We propose a complete model which allows for the simultaneous determination of whether or not to own, and how much to spend. We also suggest a maximum likelihood ratio test for the existence of simultaneity.

The paper is organized as follows. In section 2, we propose the general model and present two consistent "two stage" estimation procedures. In section 3, we point out that several limited dependent variable models can be regarded as special cases of our general model. Section 4 contains a discussion of the maximum likelihood procedure. In section 5, we propose a housing expenditure model with interdependent choices about owning or renting. In section 6 our suggested estimation procedure is applied to the housing model and the results are presented. Section 7 contains the conclusions. Finally, in the appendix we prove that our proposed "two stage" estimates are consistent.

2. A general limited dependent variable model and two stage estimation procedures.

The model we consider has the following specification.

regime 1: $Y_{1t} = X_{1t}\beta_1 + \varepsilon_{1t}$ iff $Z_t\gamma \geq \varepsilon_t$

regime 2: $Y_{2t} = X_{2t}\beta_2 + \varepsilon_{2t}$ iff $Z_t\gamma < \varepsilon_t$

That is, given exogeneous variables X_{1t}, X_{2t} and Z_t, the observed sample Y_t is generated from the first regime if the condition $Z_t\gamma \geq \varepsilon_t$ is satisfied; otherwise it is generated from the second equation. We assume ε_{1t}, ε_{2t} and ε_t are trivariately normally distributed with zero mean and a common non-singular covariance matrix for each observation. We also assume $\mathrm{var}(\varepsilon_t) = 1.$* Finally, we assume the sample observations belonging to different regimes can be distinguished. That is, we assume sample separation is available. Furthermore, only contemporaneous correlation between the disturbances is allowed.

As contrast with the model of Goldfeld, Kelejian and Quandt [9] which assumes independence between ε_t and ε_{1t}, ε_{2t} and without sample separation, the disturbances in our model are allowed to be correlated. In fact, this interdependence and sample separation are the main features of our model.

With sample separation available, we know which regime each observation belongs to. If ε_t is correlated with ε_{1t} and ε_{2t}, a least squares procedure will give biased and inconsistent estimates. This can be easily shown since $E(\varepsilon_{1t}|Z_t\gamma \geq \varepsilon_t) \neq 0$ and (or)

---

*In general, the variance of ε_t can be any constant. However in this case, γ can only be estimated up to a positive proportion. $\mathrm{var}(\varepsilon_t) = 1$ is used as a normalization rule.

$E(\varepsilon_{2t}|Z_t\gamma < \varepsilon_t) \neq 0$. So in general, a least squares model will be inappropriate for this model. In fact, because information on sample separation is available, the likelihood function for our model will be bounded and maximum likelihood procedure is applicable. If information on sample separation were not available, the likelihood function would be unbounded. The boundedness of the likelihood function follows from the arguments in Goldfeld and Quandt [8]. However the likelihood function for our model is highly nonlinear in the parameters. For this reason, the effective way to apply the maximum likelihood procedure is to start with good initial estimates.

In this section, we propose some two stage methods of getting consistent estimates for our model. These two stage methods utilize probit analysis in the first stage and a least squares procedure in second stage. Since least squares and probit procedures are very effective, these consistent estimates are easy to compute.

We will first derive a method that divides the samples into two regimes and uses the information in each regime separately. Later on we present a method that uses all the samples to estimate both regimes together.

Denote sample separation by a dichotomous variable I_t . That is,

$$I_t = 1 \text{ if and only if } Z_t\gamma \geqq \varepsilon_t ,$$

$$I_t = 0 \text{ if and only if } Z_t\gamma < \varepsilon_t .$$

With the assumption that ε_t is normally distributed, this relation is in fact the following probit model.

$$I_t^* = Z_t\gamma - \varepsilon_t$$

with $I_t = 1 \iff I_t^* \geq 0$; $I_t = 0$ otherwise. Thus we have $I_t = F(Z_t\gamma)+U_t$ where F is cumulative distribution of the standard normal variate ε_t.

Let Σ be the covariance matrix of disturbances ε_{1t}, ε_{2t} and ε_t and denote

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{1\varepsilon} \\ \sigma_{21} & \sigma_2^2 & \sigma_{2\varepsilon} \\ \sigma_{\varepsilon 1} & \sigma_{\varepsilon 2} & 1 \end{bmatrix}$$

As shown in Appendix 1, it follows that

$$E(I_t\varepsilon_t) = \sigma_{1\varepsilon}(-f(Z_t\gamma)) ,$$

$$E((1-I_t)\varepsilon_{2t}) = \sigma_{2\varepsilon}f(Z_t\gamma)$$

and $$E(I_t\varepsilon_t)^2 = \sigma_1^2F(Z_t\gamma) - \sigma_{1\varepsilon}{}^2(Z_t\gamma)f(Z_t\gamma)$$

$$E((1-I_t)\varepsilon_{2t})^2 = \sigma_2^2(1-F(Z_t\gamma)) + \sigma_{2\varepsilon}^2(Z_t\gamma)f(Z_t\gamma)$$

where f and F are standard normal density function and distribution function respectively. It then follows that

$$E(\varepsilon_{1t}|I_t = 1) = -\sigma_{1\varepsilon} f(Z_t\gamma)/F(Z_t\gamma)$$

$$E(\varepsilon_{2t}|I_t = 0) = \sigma_{2\varepsilon} f(Z_t\gamma)/(1-F(Z_t\gamma))$$

and $$E(\varepsilon_1^2|I_t = 1) = \sigma_1^2 - \sigma_{1\varepsilon}^2(Z_t\gamma)f(Z_t\gamma)/F(Z_t\gamma)$$

$$E(\varepsilon_{1t}^2|I_t = 0) = \sigma_2^2 + \sigma_{2\varepsilon}^2(Z_t\gamma)f(Z_t\gamma)/(1-F(Z_t\gamma)).$$

Hence $$E(Y_t|I_t = 1) = X_{1t}\beta_1 - \sigma_{1\varepsilon} f(Z_t\gamma)/F(Z_t\gamma)$$

$$E(Y_t|I_t = 0) = X_{2t}\beta_2 + \sigma_{2\varepsilon} f(Z_t\gamma)/(1-F(Z_t\gamma)).$$

But instead of linear equations, we have two nonlinear equations after the nonzero means have been adjusted.

$$Y_t = X_{1t}\beta_1 - \sigma_{1\varepsilon} f(Z_t\gamma)/F(Z_t\gamma) + \eta_{1t}$$

$$Y_t = X_{2t}\beta_2 + \sigma_{2\varepsilon} f(Z_t\gamma)/(1-F(Z_t\gamma)) + \eta_{2t}$$

with $E(\eta_{1t}|I_t = 1) = 0$ and $E(\eta_{2t}|I_t = 0)$.

With these equations, the two stage estimation procedure can be applied. In the first stage, we estimate the Probit model

$$I_t = F(Z_t\gamma) + U_t$$

to get consistent estimates $\hat{\gamma}$ of γ . In the second stage, we can substitute $\hat{\gamma}$ into the two nonlinear equations.

$$Y_t = X_{1t}\beta_1 - \sigma_{1\varepsilon} f(Z_t\hat{\gamma})/F(Z_t\hat{\gamma}) + \tilde{\eta}_{1t}$$

$$Y_{2t} = X_{2t}\beta_2 + \sigma_{2\varepsilon} f(Z_t\hat{\gamma})/(1-F(Z_t\hat{\gamma})) + \tilde{\eta}_{2t}.$$

With sample observations from the first regime, we can estimate β_1 and $\sigma_{1\varepsilon}$ by ordinary least squares. Similarly, we can estimate β_2 and $\sigma_{2\varepsilon}$ with sample observations from the second regime. The resultant estimates are consistent. The rigorous proofs are in Appendix 2.

The two stage estimates can be shown to be asymptotically normal. However, its asymptotical variances are relatively complicated. To improve efficiency maximum likelihood procedures can be used. Since maximum likelihood procedures require initial estimates for all its parameters, the variances and covariances have to be estimated. The estimation of σ_1^2, ${\sigma_{1\varepsilon}}^2$, σ_2^2 and ${\sigma_{2\varepsilon}}^2$ can be done by using second moments of the estimated residuals. For the observations corresponding to first regime, let

$$\hat{\varepsilon}_{1t} = Y_t - X_{1t}\hat{\beta}_1$$

and for the second regime, $\hat{\varepsilon}_{2t} = Y_t - X_{2t}\hat{\beta}_2$.

Since

$$E({\varepsilon_{1t}}^2|I_t = 1) = \sigma_1^2 - {\sigma_{1\varepsilon}}^2(Z_t\gamma)f(Z_t\gamma)/F(Z_t\gamma)$$

$$\text{and} \quad E({\varepsilon_{2t}}^2|I_t = 0) = \sigma_2^2 + {\sigma_{2\varepsilon}}^2(Z_t\gamma)f(Z_t\gamma)/(1-F(Z_t\gamma)),$$

we can use these relationships to estimate the parameters. More formally,

$$\hat{\varepsilon}_{1t}^{\,2} = \sigma_1^2 - {\sigma_{1\varepsilon}}^2(Z_t\hat{\gamma})f(Z_t\hat{\gamma})/F(Z_t\hat{\gamma}) + \zeta_{1t}$$

$$\hat{\varepsilon}_{2t}^{\,2} = \sigma_2^2 + {\sigma_{2\varepsilon}}^2(Z_t\hat{\gamma})f(Z_t\hat{\gamma})/(1-F(Z_t\hat{\gamma})) + \zeta_{2t}$$

and hence σ_1^2, $\sigma_{1\varepsilon}{}^2$, σ_2^2 can be estimated by OLS. Since estimates of $\sigma_{1\varepsilon}$ and $\sigma_{2\varepsilon}$ are obtained when we estimate β_1 and β_2, one can simplify the procedure and estimate σ_1^2 and σ_2^2 only. In this case,

$$\tilde{\sigma}_1^2 = \frac{1}{T_1}\, \Sigma_{t=1}^{T_1}(\hat{\varepsilon}_{1t}{}^2 + \hat{\sigma}_{1\varepsilon}{}^2(Z_t\hat{\gamma})f(Z_t\hat{\gamma})/F(Z_t\hat{\gamma})).$$

$$\tilde{\sigma}_2^2 = \frac{1}{T_2}\, \Sigma_{t=1}^{T_2}(\hat{\varepsilon}_{2t}{}^2 - \hat{\sigma}_{2\varepsilon}{}^2(Z_t\hat{\gamma})f(Z_t\hat{\gamma})/(1-F(Z_t\hat{\gamma}))$$

where T_1 and T_2 are the corresponding number of observed samples in regime 1 and 2 respectively. All these estimates are shown to be consistent in the Appendix.

Following this procedure, all the parameters can be estimated consistently except σ_{12}. σ_{12} cannot be estimated in this general model because it is not identifiable. The sample observations cannot reflect the correlation between Y_{1t} and Y_{2t} explicitly. This fact is brought out in the likelihood function shown below. σ_{12} does not appear anywhere in the likelihood function which is given by

$$L(\beta_1,\beta_2,\gamma,\sigma_1^2,\sigma_2^2, \sigma_{1\varepsilon},\sigma_{2\varepsilon}|Y_t, X_{1t},X_{2t},Z_t)$$
$$= \prod_t [\int_{-\infty}^{Z_t\gamma} g(Y_{1t} - X_{1t}\beta_1, \varepsilon_t)d\varepsilon_t]^{I_t}[\int_{Z_t\gamma}^{\infty} f(Y_{2t} - X_{2t}\beta_2, \varepsilon_t)d\varepsilon_t]^{1-I_t}$$

where g and f are the bivariate normal density functions of ε_{1t}, ε_t and ε_{2t}, ε_t respectively. For more specific models though, σ_{12} may be identifiable. One example of this is the disequilibrium model presented in the next section. There ε_t is an explicit function of ε_{1t} and ε_{2t} and σ_{12} is identifiable.

To combine the samples together to estimate the equations, one can use the D-method. As shown in Goldfeld and Quandt [7], this procedure is suggested if there is prior information that some of the coefficients in the two regimes are the same. Using the D-method we can write

$$Y_t = I_t Y_{1t} + (1-I_t)Y_{2t}$$

$$= I_t X_{1t}\beta_1 + (1-I_t)X_{2t}\beta_2 + I_t\varepsilon_{1t} + (1-I_t)\varepsilon_{2t} .$$

After substituting the probit equation $F(Z_t\gamma) + U_t$ for I_t and adjusting the mean of disturbance, we get

$$Y_t = F(Z_t\gamma)X_{1t}\beta_1 + (1-F(Z_t\gamma))X_{2t}\beta_2 + (\sigma_{2\varepsilon}-\sigma_{1\varepsilon})f(Z_t\gamma) + w_t$$

where $$w_t = I_t\varepsilon_{1t} + (1-I_t)\varepsilon_{2t} - (\sigma_{2\varepsilon}-\sigma_{1\varepsilon})f(Z_t\gamma) + (X_{1t}\beta_1-X_{2t}\beta_2)U_t$$

and $E(w_t) = 0$. By applying our two stage procedure to the above equation, the entire sample is used to estimate β_1 and β_2 simultaneously.

Finally, it should be noted that our two stage procedure gives not only consistent estimates but it is also a simple way to analyze identification of the model.

3. Some Special Cases of the General Model

In this section we show that other models -- a recursive model involving dichotomous endogeneous variables and some disequilibrium market models -- can be regarded as special cases of our general model. The Tobin's [23], Heckman's [10], and Nelson's [16] models can also be written in our general framework. Hence the above two stage procedures have a wide range of applicability. Recently, similar procedure has also been proposed by Heckman in the context of his model [11].

a. A Recursive Model Involving Dichotomous Endogeneous Variables

The model we will consider is

$$I^* = X_1\beta_1 - \varepsilon_1$$

$$Y_1 = X_2\beta_2 + \alpha I + \varepsilon_2$$

where I^* is an underlying and unobservable endogeneous variable. ε_1 and ε_2 are assumed to be bivariately normally distributed with zero mean and $\mathrm{var}(\varepsilon_1) = 1$. I is a dichotomous endogeneous variable which is determined by I^*. $I_1 = 1$ iff $I^* \geq 0$, otherwise $I_1 = 0$. This model is a special case of the general model since we can rewrite it as follows,

$$Y_1 = X_2\beta_2 + \alpha + \varepsilon_2 \quad \text{iff} \quad X_1\beta_1 \geq \varepsilon_1$$

$$Y_1 = X_2\beta_2 + \varepsilon_2 \quad \text{iff} \quad X_1\beta_1 < \varepsilon_1$$

Thus the model can be estimated by two stage method and the D-method is the natural way to combine the two separate equations. Firstly, the coefficients β_1 of the equation $I^* = X_1\beta_1 - \varepsilon_1$ can be estimated by Probit analysis. Substituting $I = F(X_1\hat{\beta}_1) + U_1$ into the equation for Y_1, we have

$$Y_1 = X_2\beta_2 + \alpha F(X_1\hat{\beta}_1) + w$$

In the second stage one estimates β_2 and α by ordinary least squares More detail studies on recursive models with endogenous qualitative variables will appear elsewhere.

b. Estimation of Some Disequilibrium Market Models.

Maddala and Nelson [14] and Goldfeld and Quandt [8] considered maximum likelihood estimation of four different disequilibrium market models. Here we would like to show that if information about sample separation is available, our two stage method can be applied to these disequilibrium market models. To illustrate this point, let us consider one of their models. The disequilibrium market model we consider has the following specification

$$D_t = X_{1t}\beta_1 + \varepsilon_{1t}$$

$$S_t = X_{2t}\beta_1 + \varepsilon_{2t}$$

The first equation is the demand function and the second one is the supply function. In this model, D_t and S_t cannot be observed simultaneously for each observation t , but rather we can observe the minimum quantity between them. That is,

$$Q_t = \min(D_t, S_t) .$$

The information on sample separation is assumed to be available for certain specifications. In the disequilibrium market model considered by Fair and Jaffee [5] under the title "Directional Method 1" and by Maddala and Nelson [14] as Model II, this information on sample separation is from price movement. They specify

$$\Delta P_t = P_t - P_{t-1} > 0 \quad \text{iff} \quad D_t > S_t$$

$$\Delta P_t = P_t - P_{t-1} \leq 0 \quad \text{iff} \quad S_t \geq D_t$$

and ΔP_t is observable.

This model can be rewritten as

$$Q_t = X_{1t}\beta_1 + \varepsilon_{1t} \qquad \text{iff} \quad S_t \geq D_t$$

$$\text{iff} \quad X_{2t}\beta_2 - X_{1t}\beta_1 \geq \varepsilon_{1t} - \varepsilon_{2t}$$

and

$$Q_t = X_{2t}\beta_2 + \varepsilon_{2t} \qquad \text{iff} \quad D_t > S_t$$

$$\text{iff} \quad X_{2t}\beta_2 - X_{1t}\beta_1 < \varepsilon_{1t} - \varepsilon_{2t}$$

Now, define a dichotomous variable I_t such that

$$I_t = 1 \quad \text{iff} \quad \Delta P_t \leq 0$$

$$I_t = 0 \quad \text{iff} \quad \Delta P_t > 0$$

Since ε_{1t} and ε_{2t} are assumed to be normally distributed, the following probit model can be estimated.

$$I^*_t = X_{2t}\frac{\beta_2}{\sigma^*} - X_{1t}\frac{\beta_1}{\sigma^*} - \varepsilon_t$$

where $\varepsilon_t = \frac{1}{\sigma^*}(\varepsilon_{1t} - \varepsilon_{t2})$, $\sigma^{*2} = \text{var}(\varepsilon_{1t} - \varepsilon_{t2})$ and I_t^* has the dichotomous realization I_t. Thus, this model is a special case of the general model and our two stage estimation procedure can be applied.

In contrast to the general model, the covariance of ε_{1t} and ε_{2t} can be identifiable in this model because their relations are expressed in the underlying conditions. To see this, let the covariance matrix of $(\varepsilon_{1t}, \varepsilon_{2t})$ be

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix}$$

As pointed out in the general model, we can estimate σ_1^2, σ_2^2, $\sigma_{1\varepsilon}$ and $\sigma_{2\varepsilon}$ consistently. For the disequilibrium market model

$$\sigma_{1\varepsilon} = \frac{1}{\sigma^*}(\sigma_1^2 - \sigma_{12}) \quad \text{and} \quad \sigma_{2\varepsilon} = \frac{1}{\sigma^*}(\sigma_{12} - \sigma_2^2)$$

Solving the two above equations for σ_{12} we get

$$\sigma_{12} = (\sigma_1^2\sigma_{2\varepsilon} + \sigma_2^2\,\sigma_{1\varepsilon})/(\sigma_{1\varepsilon} + \sigma_{2\varepsilon}).$$

So based on the consistent estimates of σ_1^2, σ_2^2, $\sigma_{1\varepsilon}$ and $\sigma_{2\varepsilon}$, σ_{12} can be estimated consistently.

Other disequilibrium market model such as Goldfeld and Quandt's model with sample separation [8] can also be estimated by two stage methods.

4. Maximum Likelihood Estimation

With estimates from the two stage procedures, we can use them as initial estimates for the maximum likelihood procedure. The maximum likelihood estimates in this model can easily be shown to be consistent and asymptotically efficient following Amemiya's proof [1]. Let f_1 and f_2 be the jointly normal distributions of $(\varepsilon_1, \varepsilon)$ and $(\varepsilon_2, \varepsilon)$ respectively. Dropping the t subscripts on Y, X_1, X_2, Z and I, the likelihood function for this model is

$$L(\beta_1,\beta_2,\gamma,\sigma_1^2,\sigma_2^2,\sigma_{1\varepsilon},\sigma_{2\varepsilon})$$

$$= \prod_{t=1}^{T} [\int_{-\infty}^{Z\gamma} f_1(Y - X_1\beta_1,\ \varepsilon)d\varepsilon]^I\ [\int_{Z\gamma}^{\infty} f_2(Y-X_2\beta_2,\ \varepsilon)d\varepsilon]^{1-I}$$

$$= \prod_{t=1}^{T} [\int_{-\infty}^{Z\gamma} f_1(\varepsilon|Y-X_1\beta_1)g_1(Y-X_1\beta_1)d\varepsilon]^I[\int_{Z\gamma}^{\infty} f_2(\varepsilon|Y-X_2\beta_2)g_2(Y-X_2\beta_2)d\varepsilon]^{1-I}$$

$$= \prod_{t=1}^{T} [\int_{-\infty}^{Z\gamma} f_1(\varepsilon|Y-X_1\beta_1)g_1(Y-X_1\beta_1)d\varepsilon]^I[\int_{Z\gamma}^{\infty} f_2(\varepsilon|Y-X_2\beta_2)g_2(Y-X_2\beta_2)d\varepsilon]^{1-I}$$

$$= \prod_{t=1}^{T} [g_1(Y-X_1\beta_1)\int_{-\infty}^{Z\gamma} f_1(\varepsilon|Y-X_1\beta_1)d\varepsilon]^I[g_2(Y-X_2\beta_2)\int_{Z\gamma}^{\infty} f_2(\varepsilon|Y-X_2\beta_2)d\varepsilon]^{1-I}$$

where $$g_1(Y-X_1\beta_1) = \frac{1}{\sqrt{2\pi}\ \sigma_1}\ \exp\{-\frac{1}{2\sigma_1^2}\ (Y-X_1\beta_1)^2\}$$

$$g_2(Y-X_2\beta_2) = \frac{1}{\sqrt{2\pi}\ \sigma_2}\ \exp\{-\frac{1}{2\sigma_2^2}\ (Y-X_2\beta_2)^2\}$$

$$f_1(\varepsilon|Y-X_1\beta_1) = \frac{1}{\sqrt{2\pi}\ \sqrt{1-\rho_{1\varepsilon}^2}}\ \exp\{-\frac{1}{2(1-\rho_{1\varepsilon}^2)}\ [\varepsilon-\rho_{1\varepsilon}\frac{1}{\sigma_1}(Y-X_1\beta_1)]^2\}$$

and $$f_2(\varepsilon|Y-X_2\beta_2) = \frac{1}{\sqrt{2\pi}\ \sqrt{1-\rho_{2\varepsilon}^2}}\ \exp\{-\frac{1}{2(1-\rho_{2\varepsilon}^2)}[\varepsilon-\rho_{2\varepsilon}\ \frac{1}{\sigma_2}\ (Y-X_2\beta_2)]^2\}$$

with $\rho_{1\varepsilon}$ and $\rho_{2\varepsilon}$ the correlation coefficients of $(\varepsilon_1,\varepsilon)$ and $(\varepsilon_2,\varepsilon)$ respectively.

After simplifying the integrands by transforming variables and taking logarithms, we get

$$L^* = \ln L(\beta_1,\beta_2,\gamma,\sigma_1,\sigma_2,\rho_{1\varepsilon},\rho_{2\varepsilon})$$

$$= \sum_{t=1}^{T} \{ I[\ln g_1(Y-X_1\beta_1) + \ln\int_{-\infty}^{\mu_1}\phi(t)dt] + (1-I)[\ln g_2(Y-X_2\beta_2)$$

$$+ \ln \int_{\mu_2}^{\infty}\phi(t)dt]\}$$

where $\phi(t)$ is the standard normal density function,

$$\mu_1(\beta_1,\gamma,\sigma_1,\rho_{1\varepsilon}) = \frac{1}{\sqrt{1-\rho_{1\varepsilon}^2}}[Z\gamma- \frac{\rho_{1\varepsilon}}{\sigma_1}(Y-X_1\beta_1)]$$

and $$\mu_2(\beta_2,\gamma,\sigma_2,\rho_{2\varepsilon}) = \frac{1}{\sqrt{1-\rho_{2\varepsilon}^2}}[Z\gamma- \frac{\rho_{2\varepsilon}}{\sigma_2}(Y-X_2\beta_2)]$$

To calculate the maximum likelihood estimates numerical optimization algorithms must be used. A survey on different algorithms can be found in Goldfeld and Quandt [7]. Since the likelihood function derived above is highly nonlinear, these computations will be quite complicated. However because our initial estimates are consistent, a simpler linear version of the maximum likelihood technique can be used. These estimates are obtained by using one iteration of the Newton-Raphson method. Let $\tilde{\Theta}$ be the consistent estimates of Θ, where $\Theta' = (\beta_1', \beta_2',\gamma',\sigma_1,\sigma_2,\rho_{1\varepsilon},\rho_{2\varepsilon})$.

The two step maximum likelihood estimates (2SML) $\bar{\Theta}_M$ are calculated from

$$\bar{\Theta}_M = \tilde{\Theta} -[\frac{\partial^2 \ln L(\tilde{\Theta})}{\partial\Theta\, \partial\Theta'}]^{-1} \frac{\partial \ln L(\tilde{\Theta})}{2\Theta}$$

$\tilde{\theta}_M$ is consistent, asymptotically efficient and asymptotically normal. The proof of this is similar to the proof found in Amemiya [1]. The estimated standard errors of the maximum likelihood estimates are the square roots of the diagonal elements of

$$-\left[\frac{\partial^2 \ell nL(\tilde{\theta})}{\partial\theta\, \partial\theta'}\right]^{-1}$$

Hence, statistical tests of significance can be easily performed.

The expressions for these second derivatives are relatively complicated but they can be derived straightforwardly from the first derivatives. The first derivatives for this likelihood function are shown below.

$$\frac{\partial L^*}{\partial\beta_1} = \sum_{t=1}^{T} I\left[\frac{(Y-X_1\beta_1)}{\sigma_1^2} X_1' + \frac{\phi(\mu_1)\rho_{1\varepsilon}X_1'}{\sigma_1\sqrt{1-\rho_{1\varepsilon}^2}\; F(\mu_1)}\right]$$

$$\frac{\partial L^*}{\partial\beta_2} = \sum_{t=1}^{T} (1-I)\left[\frac{(Y-X_2\beta_2)}{\sigma_2^2} X_1' - \frac{\phi(\mu_2)\rho_{2\varepsilon}X_2'}{\sigma_2\sqrt{1-\rho_{2\varepsilon}^2}\; (1-F(\mu_2))}\right]$$

$$\frac{\partial L^*}{\partial\gamma} = \sum_{t=1}^{T} \left\{I \frac{\phi(\mu_1)Z'}{\sqrt{1-\rho_{1\varepsilon}^2}\; F(\mu_2)} - (1-I)\frac{\phi(\mu_2)Z'}{\sqrt{1-\rho_{2\varepsilon}^2}\; (1-F(\mu_2))}\right\}$$

$$\frac{\partial L^*}{\partial\sigma_1} = \sum_{t=1}^{T} I\left[-\frac{1}{\sigma_1} + \frac{1}{\sigma_1^3}(Y-X_1\beta_1)^2 + \frac{\rho_{1\varepsilon}\phi(\mu_1)(Y-X_1\beta_1)}{\sigma_1^2\sqrt{1-\rho_{1\varepsilon}^2}\; F(\mu_1)}\right]$$

$$\frac{\partial L^*}{\partial \sigma_2} = \sum_{t=1}^{T} (1-I)\left[-\frac{1}{\sigma_2} + \frac{1}{\sigma_2^3}(Y-X_2\beta_2)^2 - \frac{\rho_{2\varepsilon}\phi(\mu_2)(Y-X_2\beta_2)}{\sigma_2^2\sqrt{1-\rho_{2\varepsilon}^2}\,(1-F(\mu_2))}\right]$$

$$\frac{\partial L^*}{\partial \rho_{1\varepsilon}} = \sum_{t=1}^{T} I\,\frac{\phi(\mu_1)(\rho_{1\varepsilon}\mu_1 - \sqrt{1-\rho_{1\varepsilon}^2}\,(\frac{Y-X_1\beta_1}{\sigma_1}))}{F(\mu_1)(1-\rho_{1\varepsilon}^2)}$$

$$\frac{\partial L^*}{\partial \rho_{2\varepsilon}} = \sum_{t-1}^{T} (1-I)\frac{\phi(\mu_2)(\sqrt{1-\rho_{2\varepsilon}^2}\,(\frac{Y-X_2\beta_2}{\sigma_2}) - \rho_{2\varepsilon}\mu_2)}{(1-F(\mu_2))(1-\rho_{2\varepsilon}^2)}$$

where ϕ and F are the standard normal density and distribution functions respectively.

5. A Housing Demand Model and Interdependent Choices - An Application

Several facets of government policy are aimed at the problem of providing adequate housing for lower income families. Efforts such as public housing and FHA - subsidized mortgages are directed towards the supply side, while other policies such as housing subsidies and transfer payments are directed towards the demand side. To aid government in these policy decisions many models of housing demand have been presented in the literature. One approach is to use ordinary least squares estimation techniques and estimate demand or expenditure equations. In this approach the sample data are usually divided into owners and renters and demand equations are estimated separately for each group. Alternatively the entire sample can be used to estimate a single equation with several dummy variables. For examples of the least squares approach we can see Reid [21], Lee [13] or Carliner [4]. Another approach models only the choice of whether to buy or rent. This can be accomplished with a linear probability model, a Probit model or a logit model. With this method one does not obtain demand equations but rather an equation to predict the probability that a given family will own their own home. Examples of the Probit model can be found in Ohls [17] and Poirier [18]. Quigley [20], in a similar analysis, uses the Logit model to estimate the probability that a given family will choose among 18 types of residential housing.

But a typical sample is comprised of housing expenditures by families who either own their own home or rent. A complete model then should take into account the joint determination of whether or not to own and how much to spend. In the two above approaches only one part of this complete model is emphasized. Furthermore, ordinary least squares will give biased estimates if there is simultaneity between the expenditures and choice equations. One may argue that this simultaneity does not occur and hence ordinary least squares is valid. However, before drawing this conclusion a rigorous statistical test should be performed.

To formulate the joint determination of whether or not to own and how much to spend, a model similar to the one specified in section 2 is used. More specifically, the model we study has the following specification

$$C_{1t} = X_{1t}\beta_1 + \varepsilon_{1t} ,$$

$$C_{2t} = X_{2t}\beta_2 + \varepsilon_{2t} ,$$

$$I_t^* = Z_t\gamma - \varepsilon_t$$

where C_{1t} is annual expenditures on housing if the family owns. C_{2t} is annual expenditures on housing if the family rents, and I_t^* is an unobservable index which determines the choices. X_{1t} X_{2t} and Z_t are vectors of exogeneous variables. The assumptions about the disturbances are the same as the ones in our general model. Since C_{1t} and C_{2t} are assumed to be mutually exclusive, they cannot be observed simultaneously for any one individual. What we do observe are the binary index I_t, the families' decisions to own or rent, and expenditures on housing C_t such that

$$C_t = C_{1t} ; \quad I_t = 1 \quad \text{iff} \quad I^* \geq 0$$

$$C_t = C_{2t} ; \quad I_t = 0 \quad \text{otherwise}$$

The exogeneous variables X_{1t}, X_{2t} and Z_t are always observable.

To study this model we use one set of aggregate price indices for both owners and renters. Each household has associated with it a housing price index and a price index for all goods. Given this price specification the relative price of housing services would not enter into the rent-own decision function. It would, of course, influence the allocation of the budget between housing and all other goods. The same explanatory variables are used in the two expenditure equations. They include the personal characteristics of family head (age, race, sex); family background (mover, family "permanent income", family size), regional variables (city size and distance from the center of the city) and a relative price index of housing. All these explanatory variables except the relative price of housing are included in the decision function.

The price indices were constructed from table 128 in the Bureau of Labor Statistics (BLS) Handbook of Labor Statistics 1972 - "Consumer Price Index, 23 Cities or Standard Metropolitan Statistical Areas. All Items and Major Groups, 1947-1971". This table gives six price indices -- including a housing index -- with a base of 1967 = 100. Twenty-one of these SMSA's (Honolulu and Washington, D.C. were excluded) were grouped into one of four regions: Northeast, North Central, South and West. Two average price indices and a relative price index were then calculated for each of the four regions and assigned to the appropriate family. The rest of the data are from A Panel Study of Income Dynamics [22] on a sample of 3028 families in 1971 (henceforth, the "Panel").

The annual expenditure variables were housing cost divided by a price of housing index. The housing cost variable is described in the Panel. It includes utility payments, value of additions and repairs done by the family, property taxes for homeowners, 6% of house value for homeowners and annual rent payments for families who rent. That is, the dependent variable in our housing demand equations is units of housing as measured in real annual dollar expenditures. The permanent income variable used was a five year average of family incomes plus imputed rental income for homeowners (i.e., 6% of net equity), all divided by a price index. The family income variable includes labor income of head and/or wife; asset income from farm or business; rental, interest and dividend income; and transfer payments such as Aid to Dependent Children. The variables of city population and distance from the Center of the nearest city of 50,000 population or more were included to capture regional price differences as might be caused by different supply and demand conditions in different locations. The family size variable was included to capture demographic differences. This variable is defined as number of people (children plus adults) living in the family unit. The mover variable is a dummy variable if the family moved more than once between 1968 and 1972. The reason for including this variable in the demand equations was to capture the different search costs as well as different demographic factors of the mobile families. It was included in the decision function I^* because of the transactions cost involved in buying a home. The moving family is more likely to rent than own. Finally, the dummies for age, sex and race of head were included to capture the differences in tastes and expenditure patterns for different groups of individuals. For empirical purposes, the continuous variables were measured in logarithmic scale. This makes the estimated demand equations more directly comparable to previous studies.

6. Empirical Results and Tests for Simultaneity

In Tables 1 and 2 we present the probit and two step maximum likelihood (2SML) estimates of the decision function respectively. Both estimation procedures yield similar results, but the 2SML estimates have slightly smaller standard errors for all the estimated parameters. Both age dummies are significant and indicate that families with a head over 64 are the most likely to own their own home and as expected, families headed by an individual under 36 are the least likely to own. The coefficient for the Black dummy (1 for black) is highly significant and negative, indicating that Blacks are more likely to be renters than owners. This could be due to either price discrimination or merely a stronger preference for renting among Blacks. Similarly, the negative sign for the female dummy indicates either market discrimination or a stronger preference for renting among females. The negative coefficient for the mover variable suggests that transaction costs of buying and selling a house effectively raise the price of a home for the family that frequently moves. These families are therefore more likely to rent. The next two variables, city size and distance from the center of the city, were included to capture supply and demand differences in different locations not captured by the price variables. The coefficients indicate that in rural areas the family

---

*In the Panel these two variables were coded in groups. For example city size was classified as being in one of six groups: greater than 500,000; 100,000 to 499,999, etc. To make the variable continuous a real number, the midpoint when appropriate, was used for each group. That is, for the two above examples the family was assigned a value of 1,000,000 or 300,000 respectively.

is more likely to own their own home. The coefficient for family size is positive and this indicates that large families prefer the more spacious living conditions provided by homeownership. The positive coefficient for the income variable indicates that high income families are more likely to own their own homes. This may be due to the fact that lower income families are unable to obtain a mortgage for the size of house they desire and therefore rent rather than own. It should be noted that all the coefficients are significant and hence they have significant impact on the buy vs. rent decision.

As we formulate our model without any equality constraint on coefficients in the two expenditure equation, the consistent two stage estimates are obtained by splitting the sample into two groups, owners and renters. In Table 3 we present the two stage and 2SML estimates of the expenditure equation for owners. Table 4 contains estimates for renters. The two stage consistent estimates and 2SML estimates are quite similar with three exceptions. First, the mover coefficient in the owner equation changes from the two stage estimate of -0.04407 to the 2SML estimate of -0.166199. Second, the city size coefficient in the owner equation decreases from the two stage estimate of 0.06592 to the 2SML estimate of 0.04237. In terms of the 2SML standard errors, these represent more than a three standard error change. Third, the distance variable in the renter equation changes from a positive sign to a negative sign and is significant. With the exception of price in the renter equation, all the 2SML coefficients are significant.

TABLE 1: Probit Estimates of the Decision Function

VARIABLES	ESTIMATES	STANDARD ERROR
Constant	-3.26292	0.47282
Age $\leq$ 35 dummy	-1.17045	0.10754
36 $\leq$ age $\leq$ 64	-0.63406	0.09768
black dummy	-0.30715	0.07134
female dummy	-0.32102	0.06729
mover dummy	-0.94359	0.07093
ℓn (city size)	-0.14969	0.01763
ℓn (distance from center of city)	0.12779	0.02721
ℓn (family size)	0.13434	0.04929
ℓn (relative permanent income)	0.68437	0.05168

TABLE 2: Two Steps Maximum Likelihood Estimates of the Decision Function

VARIABLES	ESTIMATES	STANDARD ERROR
Constant	-3.24123	0.47183
age $\leq$ 35 dummy	-1.16230	0.10681
36 $\leq$ age $\leq$ 64 dummy	-0.63151	0.09711
black dummy	-0.30971	0.07076
female dummy	-0.31895	0.06685
mover dummy	-0.94527	0.07053
ℓn (city size)	-0.14991	0.01755
ℓn (distance from center of city)	0.12864	0.02700
ℓn (family size)	0.13381	0.04904
ℓn (relative permanent income)	0.68159	0.05155

The coefficients for the age dummies are negative. This implies that families with a head over 64 have larger real dollar expenditures on housing than other age groups. Blacks may be constrained to low quality neighborhoods and thus have a stronger preference for other goods. Hence the coefficient for black dummy is negative. The positive coefficient for females suggests that females have a stronger preference for housing than males. While Table 1 suggested that females may be discriminated against in the mortgage market, Table 3 indicates that they tend to spend more than males if they own houses. The estimated signs of mover variables are different for the two groups. The coefficient in the renter group is positive. This may be the result of their search strategies for rental prices. That is, since they spend relatively short time in each location, the marginal benefit from an additional search does not outweigh the marginal cost. For the nonmoving renter the optimal searching rule probalby results in a relatively lower expenditure. On the other hand, the mover coefficient was negative in the owner group. This may be due to several factors. First, transaction costs are lower for less expensive houses since realtor fees and the like are a percentage of the cost. Mobile families can economize on these costs by buying less expensive dwellings. Second, it may be easier to resell a lower priced house. Third, mobile families may have little incentive to maintain their houses. Rather, they are more likely to settle for a simple place to live. The estimated signs of the city size and distance from center of city coefficients

TABLE 3: Housing Expenditure Equation of the Owners(*)

VARIABLES	CONSISTENT ESTIMATES BY TWO STAGE METHOD	TWO STEPS MAXIMUM LIKELIHOOD ESTIMATES	
Constant	1.71337	1.53854	(0.19282)
age ≤ 35	-0.09893	-0.20352	(0.04034)
36 ≤ age ≤ 64	-0.13742	-0.19019	(0.03331)
black	-0.2446	-0.29840	(0.03215)
female	0.10597	0.06638	(0.03109)
mover	-0.04407	-0.16199	(0.03608)
ℓn (city size)	0.06592	0.04237	(0.00671)
ℓn (distance from center of city)	-0.01036	-0.02542	(0.01140)
ℓn (family size)	0.03032	0.0456	(0.02268)
ℓn (relative price of housing)	-2.644	-2.58558	(1.07473)
ℓn (real permanent income)	0.55365	0.60678	(0.02008)

$\hat{\sigma}_1$ (standard error of disturbances in expenditure equation, max. likelihood estimate) = 0.43934

$\hat{\rho}_{1\varepsilon}$ (correlation coefficient of disturbances between expenditures equation and decision function, max. likelihood estimate) = - 0.21629

(*) The numbers in brackets are standard errors.

TABLE 4: Housing Expenditure Equation of the Renters(*)

VARIABLES	CONSISTENT ESTIMATES BY TWO STAGE METHOD	TWO STEPS MAXIMUM LIKELIHOOD ESTIMATES	
One	1.77068	1.96969	(0.23043)
Age $\leq$ 35	-0.18588	-0.18420	(0.05722)
36 $\leq$ age $\leq$ 64	-0.12270	-0.12180	(0.05571)
black	-0.19740	-0.19897	(0.03119)
female	0.11713	0.11444	(0.02818)
mover	0.13769	0.13453	(0.02782)
ℓn (city size)	0.06946	0.06145	(0.00854)
ℓn (distance from center of city)	0.00716	-0.02581	(0.012221)
ℓn (family size)	0.08324	0.08666	(0.02035)
ℓn (relative price of housing)	-1.81520	-2.11270	(1.37482)
ℓn (real permanent income)	0.50298	0.50046	(0.02324)

$\hat{\sigma}_2$ (maximum likelihood estimate) = 0.4307

$\hat{\rho}_{2\epsilon}$ (correlation coefficient of the disturbances between expenditure equation and decision function, maximum likelihood estimate) = -0.00473

(*) The numbers in brackets are standard errors.

are compatible. People spend more on housing as they live near the city and as the city sizes increases. As mentioned earlier, these variables may be capturing price differences. The signs are expected as housing is more expensive in large city and urban areas. The positive sign for the family size variable is expected. The larger the family, the larger is the required living space. Hence, large families have to spend relatively more on housing than small families. Although past papers such as Lee [13], Carliner [4], Fenton [6], and Polinsky [19] specified slightly different expenditure equations than ours, the estimates of income elasticity are well below one and similar to ours. Finally, the coefficients for the relative price variables are negative. The negative sign does indicate that families tend to spend less on housing in those regions where housing is relatively (to all other goods) more expensive.

To compare our estimates of the expenditure equations with OLS estimates, the OLS estimates were obtained. These estimates are presented in Table 5 and 6. A comparison of Table 3 and 5 shows the 2SML and OLS estimates to be slightly different but compatible. One exception to this compatibility is the distance variable coefficient. The value obtained with 2SML is five times greater than the OLS estimate. Also, the 2SML estimate is significant and the OLS estimate is insignificant. By comparing Tables 4 and 6 we again see a difference in the distance coefficient. The 2SML estimate is negative and significant. The OLS estimate is positive and insignificant.

TABLE 5: Housing Expenditure Equation of the Owners

VARIABLES	O.L.S. ESTIMATES	STANDARD ERROR
constant	1.49477	
age $\leq$ 35	-0.15345	0.0413
36 $\leq$ age $\leq$ 64	-0.16520	0.03359
black	-0.26055	0.05709
female	0.08828	0.03129
mover	-0.09916	0.03732
ℓn (city size)	0.05922	0.00572
ℓn (distance from center of city)	-0.00563	0.03646
ℓn (family size)	0.03752	0.02284
ℓn (relative price of housing)	-2.62209	1.085
ℓn (real permanent income)	0.58399	0.02079

Estimated standard error of distances $\hat{\sigma}_1 = 0.44996$

$R^2 = 0.5188$

TABLE 6: Housing Expenditure Equation of the Renters

VARIABLES	O.L.S. ESTIMATES	STANDARD ERROR
constant	1.77267	
age $\leq$ 35	-0.18423	0.05846
36 $\leq$ age $\leq$ 64	-0.12172	0.05634
black	-0.19701	0.0465
female	0.11751	0.02836
mover	0.13878	0.02877
ℓn (city size)	0.06964	0.00778
ℓn (distance from center of city)	0.00698	0.02998
ℓn (family size)	0.08309	0.0205
ℓn (relative price of housing)	-1.81576	1.36963
ℓn (real permanent income)	0.50212	0.02368

Estimated standard error of disturbance $\hat{\sigma}_2 = 0.4348$ $R^2 = 0.45078$

While these casual comparisons do reveal some differences in the two approaches, a more detailed investigation on whether or not simultaneity occurs is necessary. To carry out this investigation, we use a maximum likelihood ratio test. Consider the null hypothesis that there is no correlation between the disturbances in the housing expenditure equations and the decision function. That is, the null hypothesis constrains $\sigma_{1\varepsilon}$ and $\sigma_{2\varepsilon}$ to zero. The maximum likelihood estimates of the expenditure equations now reduce to simple OLS estimates and the maximum likelihood estimates of the decision function are the probit estimates. Denote these estimates as $\hat{\Theta}_o$. The alternative hypothesis does not constrain $\sigma_{1\varepsilon}$ and $\sigma_{2\varepsilon}$ to zero. The maximum likelihood estimates are now the 2SML estimates presented in Table 2-4. Denote these estimates as $\hat{\Theta}_M$. With $\hat{\Theta}_o$ and $\hat{\Theta}_M$ the ratio $-2\ell n \frac{L(\hat{\Theta}_o)}{L(\hat{\Theta}_M)}$ has a χ^2_2 distribution asymptotically. For our model, this ratio is 6.771 and it is significant at the 0.05 level. This implies that simultaneity does occur, albeit the evidence is weak. For this model we conslude that OLS yields biased estimates of the expenditure equations, even if this bias is not very large. Therefore estimation procedures such as ours, which account for the simultaneity are more appropriate. For this model, we have also utilized the D-method in the two stage estimation but it is less satisfactory than using the separate samples. In fact, it gives a smaller likelihood value. This implies that it is more appropriate to use the samples separately.

7. Conclusions

In this paper we study a general limited dependent variables model and point out that many of the limited dependent variable models presented in the literature can be regarded as special cases of our model. We derive simple methods to get consistent estimates for our model. These are two stage methods. The first stage is probit analysis and the second stage is ordinary least squares. The two stage estimates are proved to be consistent under general conditions. These methods give not only consistent estimates but are easy ways to analyze the identification of the model. With these consistent estimates, a two step maximum likelihood method which gives asymptotic efficient estimates is then derived.

Our estimation procedures are used to study a housing expenditure model. This model differs from previous studies in that it takes into account the simultaneous determination of how much to spend and whether or not to own or rent. Our estimation procedures proved practical even though our model has a large number of unknown parameters and is estimated from a large sample. Also the procedures perform quite well in terms of standard errors and explained variation. The empirical results are quite compatible with economic theory.

Finally, to test the hypothesis that simultaneity does exist in the housing model, we compare our 2SML estimates to the usual OLS estimates. By using a maximum likelihood ratio test we found evidence that simultaneity does exist. Therefore we can conclude that our approach to estimate the expenditure equations is more appropriate than the usual OLS approach.

Appendix 1: Truncated Mean and Variance

In this section, we derive the formulae for the mean and variances used in the previous sections.

Proposition: Let $(\varepsilon_{1t}, \varepsilon_t) \sim N(0, \begin{bmatrix} \sigma_1^2 & \sigma_{1\varepsilon} \\ \sigma_{1\varepsilon} & 1 \end{bmatrix})$ and I_t is a dichotomous variable which is defined as

$$I_t = 1 \text{ if and only if } Z_t\gamma \geqq \varepsilon_t \,,$$

$$I_t = 0 \text{ if and only if } Z_t\gamma < \varepsilon_t \,.$$

Then $E(I_t\varepsilon_{1t}) = \sigma_{1\varepsilon}(-\frac{1}{\sqrt{2\pi}} e^{-(Z_t\gamma)^2/2})$,

$$E[(I_t\varepsilon_{1t})^2] = \sigma_1^2 F(Z_t\gamma) - \sigma_{1\varepsilon}^2 (Z_t\gamma)(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(Z_t\gamma)^2})$$

Proof:

$$E(I_t\varepsilon_{1t}) = E(\varepsilon_{1t} | I_t = 1) \quad P(I_t = 1)$$

$$= (\int_{-\infty}^{\infty} \varepsilon_{1t} g(\varepsilon_{1t} | I_t = 1) \, d\varepsilon_{1t}) P(I_t = 1)$$

where $g(\varepsilon_{1t} | I_t = 1)$ denotes the conditional density of ε_{1t} conditioned on $I_t = 1$. It implies

$$E(I_t\varepsilon_{1t}) = \int_{-\infty}^{\infty} \varepsilon_{1t} (\int_{-\infty}^{Z_t\gamma} g(\varepsilon_{1t}, \varepsilon_t) d\varepsilon_t) d\varepsilon_{1t}$$

where $g(\varepsilon_{1t}, \varepsilon_t)$ is the bivariate normal density function of ε_{1t} and ε_t. It then follows that

$$E(I_t\varepsilon_{1t}) = \int_{-\infty}^{Z_t\gamma} (\int_{-\infty}^{\infty} \varepsilon_{1t} g(\varepsilon_{1t} | \varepsilon_t) d\varepsilon_{1t}) g(\varepsilon_t) d\varepsilon_t .$$

Since $\sigma_\varepsilon^2 = 1$, $g(\varepsilon_{1t} | \varepsilon_t) = \frac{1}{\sqrt{2\pi}\,\sigma_1 \sqrt{1-\rho_1^2}} \exp\{-\frac{1}{2\sigma_1^2 (1-\rho_1^2)} [\varepsilon_{1t} - \rho_1 \sigma_1 \varepsilon_t]^2\}$

where σ_1, ρ_1 denote the standard deviation of ε_{1t} and the correlation coefficient between ε_{1t} and ε_t . $g(\varepsilon_t)$ is the standard normal density function. Thus

$$E(I_t\varepsilon_{1t}) = \int_{-\infty}^{Z_t\gamma} (\int_{-\infty}^{\infty} \varepsilon_{1t} \frac{1}{\sqrt{2\pi}\,\sigma_1\sqrt{1-\rho_1^2}} \exp\{-\frac{1}{2\sigma_1^2(1-\rho_1^2)} [\varepsilon_{1t}-\rho_1\sigma_1\varepsilon_t]^2\}$$

$$d\varepsilon_{1t})g(\varepsilon_t)d\varepsilon_t \ .$$

Let $u_t = \varepsilon_{1t}-\rho_1\sigma_1\varepsilon_t$; we have

$$E(I_t\varepsilon_{1t}) = \int_{-\infty}^{Z_t\gamma} (\int_{-\infty}^{\infty}(u_t+\rho_1\sigma_1\varepsilon_t) \frac{1}{\sqrt{2\pi}\,\sigma_1\sqrt{1-\rho_1^2}} \exp\{\frac{1}{2\sigma_1^2(1-\rho_1^2)} u_t^2\}du_t)$$

$$g(\varepsilon_t)d\varepsilon_t$$

$$= \int_{-\infty}^{Z_t\gamma} \rho_1\sigma_1\varepsilon_t g(\varepsilon_t)d\varepsilon_t$$

$$= \rho_1\sigma_1 \int_{-\infty}^{Z_t\gamma} \varepsilon_t g(\varepsilon_t)d\varepsilon_t$$

$$= \sigma_{1\varepsilon}(-\frac{1}{\sqrt{2\pi}} e^{-(Z_t\gamma)^2/2})$$

Also

$$E(I_t\varepsilon_{1t})^2 = \int_{-\infty}^{\infty} \varepsilon_{1t}^2 (\int_{-\infty}^{Z_t\gamma} g(\varepsilon_{1t},\varepsilon_t)d\varepsilon_t)d\varepsilon_{1t}$$

$$= \int_{-\infty}^{Z_t\gamma} (\int_{-\infty}^{\infty} \varepsilon_{1t}^2\, g(\varepsilon_{1t}|\varepsilon_t)\,d\,\varepsilon_{1t})g(\varepsilon_t)d\varepsilon_t$$

$$= \int_{-\infty}^{Z_t\gamma} (\sigma_1^2(1-\rho_1{}^2) + (\rho_1\sigma_1\varepsilon_t)^2)g(\varepsilon_t)d\varepsilon_t$$

$$= \int_{-\infty}^{Z_t\gamma} [\sigma_1^2 - \sigma_{1\varepsilon}^{\ 2} + \sigma_{1\varepsilon}^{\ 2}\varepsilon_t^{\ 2}] g(\varepsilon_t) d\varepsilon_t$$

$$= (\sigma_1^2 - \sigma_{1\varepsilon}^{\ 2}) F(Z_t\gamma) + \sigma_{1\varepsilon}^{\ 2} \int_{-\infty}^{Z_t\gamma} \varepsilon_t^{\ 2} g(\varepsilon_t) d\varepsilon_t$$

$$= \sigma_1^2 F(Z_t\gamma) - \sigma_{1\varepsilon}^2 \ (Z_t\gamma)\left(\frac{1}{\sqrt{2\pi}} \ e^{-1/2(Z_t\gamma)^2}\right) \qquad \text{Q.E.D.}$$

Appendix 2: Consistency of the Two Stage Estimator

In this appendix, we would like to prove under certain conditions, the two stage estimators are consistent.

Consider the general nonlinear equation

$$Y_t = f(X_t, \beta_o)\alpha + \varepsilon_t$$

where X_t are given exogeneous variables, ε_t are serially independent and have zero mean.

$$f : X \times H \rightarrow R^k$$

where H denotes the parameter space of β and $X \subseteq R^m$. Also

$$f(X_t, \beta) = (f_1(x_t, \beta), \ldots, f_k(X_t, \beta)).$$

Let $\tilde{\beta}_T$ be a consistent estimator of β_o. The two stage estimator is defined as

$$\hat{\alpha}_T = \left(\begin{bmatrix} f_1(X_1, \tilde{\beta}_T) \ldots\ldots f_1(X_T, \tilde{\beta}_T) \\ \vdots \qquad\qquad \vdots \\ f_k(X_1, \tilde{\beta}_T) \ldots\ldots f_k(X_T, \tilde{\beta}_T) \end{bmatrix} \begin{bmatrix} f_1(X_1, \tilde{\beta}_T) \ldots\ldots f_k(X_1, \tilde{\beta}_T) \\ \vdots \qquad\qquad \vdots \\ f_1(X_T, \tilde{\beta}_T) \ldots\ldots f_k(X_T, \tilde{\beta}_T) \end{bmatrix} \right)^{-1}$$

$$\begin{bmatrix} f_1(X_1, \tilde{\beta}_T) \ldots\ldots f_1(X_1, \tilde{\beta}_T) \\ \vdots \qquad\qquad \vdots \\ f_k(X_1, \tilde{\beta}_T) \ldots\ldots f_k(X_T, \tilde{\beta}_T) \end{bmatrix} \begin{bmatrix} Y_1 \\ \vdots \\ Y_T \end{bmatrix}$$

Proposition 1. In the above nonlinear regression equation, the two stage estimator $\hat{\alpha}_T$ is strongly consistent if the following conditions are satisfied:

1). $\{\varepsilon_t\}$ is independent for different t with zero mean and its first four moments are uniformly bounded.

2). $f(X_t, \beta)$ is continuous on $X \times H$ where H is compact and β_o is an interior point in H.

3). $\{X_t\}$ is bounded, and the empirical distribution G_n defined by $G_n(X) = j/n$, where j is the number of points $x_1, x_2, \ldots, x_n$ less than or equal to x, converges to a distribution function.

4)

$$\lim_{T\to\infty} \frac{1}{T} \begin{bmatrix} \Sigma_{t=1}^{T} f_1^2(X_t, \beta_o) & \cdots\cdots\cdots & \Sigma_{t=1}^{T} f_1(X_t, \beta_o) f_k(X_t, \beta_o) \\ \vdots & & \vdots \\ \Sigma_{t=1}^{T} f_k(X_t, \beta_o) f_1(X_t, \beta_o) & \cdots & \Sigma_{t=1}^{T} f_k^{\,2}(X_t, \beta_o) \end{bmatrix}$$

is positive definite.

5) $\tilde{\beta}_T$ is a strongly consistent estimator of β_o

Proof: The proof of this proposition is based on the theorems by Jennrich [12] and Amemiya [1].

$$\hat{\alpha}_T = \begin{bmatrix} \Sigma_{t=1}^{T} f_1^{\,2}(X_t, \tilde{\beta}_T) & & \Sigma_{t=1}^{T} f_1(X_t, \tilde{\beta}_T) f_k(X_t, \tilde{\beta}_T) \\ & & \\ \Sigma_{t=1}^{T} f_k(X_t, \tilde{\beta}_T) f_1(X_t, \tilde{\beta}_T) & \cdots\cdots & \Sigma_{t=1}^{T} f_k^{\,2}(X_t, \tilde{\beta}_T) \end{bmatrix}^{-1}$$

$$\left(\begin{bmatrix} \Sigma_{t=1}^{T} f_1(X_t,\tilde{\beta}_T) f_1(X_t,\beta_o) & \cdots & \Sigma_{t=1}^{T} f_1(X_t,\tilde{\beta}_T) f_k(X_t,\beta_o) \\ \vdots & & \vdots \\ \Sigma_{t=1}^{T} f_k(X_t,\tilde{\beta}_T) f_1(X_t,\beta_o) & \cdots & \Sigma_{t=1}^{T} f_k(X_t,\tilde{\beta}_T) f_k(X_t,\beta_o) \end{bmatrix} \alpha + \begin{bmatrix} \Sigma_{t=1}^{T} f_1(X_t,\tilde{\beta}_T)\varepsilon_t \\ \vdots \\ \Sigma_{t=1}^{T} f_k(X_t,\tilde{\beta}_T)\varepsilon_t \end{bmatrix}\right)$$

By lemma 1 in Amemiya, $\forall$ i, j =1, ..., k

$$\frac{1}{T} \sum_{t=1}^{T} f_i(X_t,\beta) f_j(X_t,\beta) \quad \text{converges to} \quad \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} f_i(X_t,\beta) f_j(X_t,\beta)$$

uniformly for β in H . Thus by lemma 4,

$$\frac{1}{T} \sum_{t=1}^{T} f_i(X_t,\tilde{\beta}_T) f_j(X_t,\tilde{\beta}_T) \quad \text{converges to} \quad \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} f_i(X_t,\beta_o) f_j(X_t,\beta_o) \text{ a.e.}$$

Also by lemma 1, we have

$$\frac{1}{T} \sum_{t=1}^{T} f_i(X_t,\zeta) f_j(X_t,\beta) \quad \text{converges to} \quad \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} f_i(X_t,\zeta) f_j(X_t,\beta)$$

uniformly for ζ, β in H x H and hence

$$\frac{1}{T} \sum_{t=1}^{T} f_i(X_t,\tilde{\beta}_T) f_j(X_t,\beta_o) \text{ converges to } \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} f_i(X_t,\beta_o) f_j(X_t,\beta_o).$$

It follows $\lim_{T\to\infty} \hat{\alpha}_T = \alpha + \lim_{T\to\infty} \frac{1}{T} \begin{bmatrix} \Sigma_{t=1}^{T} f_1(X_t,\tilde{\beta}_T)\varepsilon_t \\ \Sigma_{t=1}^{T} f_k(X_t,\tilde{\beta}_T)\varepsilon_t \end{bmatrix}$

Since $\frac{1}{T} \sum_{t=1}^{T} f_i(X_t,\zeta) f_i(X_t,\beta)$ converges to $\lim_{T\to\infty} \sum_{t=1}^{T} f_i(X_t,\zeta) f_j(x_t,\beta)$ uniformly for ζ, β in $H \times H$, by lemma 2 in Amemiya, $\frac{1}{T} \sum_{t=1}^{T} f_i(X_t,\zeta)\varepsilon_t$ converges to 0 a.e. uniformly for all ζ in H. It follows again by lemma 4 that

$$\frac{1}{T}\Sigma_{t=1}^{T} f_i(X_t,\tilde{\beta}_T)\varepsilon_t \text{ converges to } 0 \text{ a.e.}$$

Hence we can conclude that $\lim_{T\to\infty} \hat{\alpha}_T = \alpha$ a.e. Q.E.D.

In the two stage estimation procedure, we also estimate, the variance parameters by using the estimated residuals. To show its consistency, let us write them down in a more general model.

Consider the regression equations

$$Y_t = f(X_t,\beta_o)\alpha_o + \varepsilon_t$$

$$\varepsilon_t^2 = g(X_t,\alpha_o,\beta_o)\gamma + U_t$$

where ε_t, u_t are serially independent and have zero means. The α_o, β_o, γ are the true parameter of the model in the parameter space H_1, H_2 and H_3 which are subsets of R^k, R^ℓ, R^M respectively.

Let $\tilde{\alpha}_T$, $\tilde{\beta}_T$ be the estimates of α_o and β_o. The two stage estimator of γ is defined as

$$\hat{\gamma}_T = \left(\begin{bmatrix} g_1(X_1,\tilde{\alpha}_T,\tilde{\beta}_T) \cdots\cdots g_1(X_T,\tilde{\alpha}_T,\tilde{\beta}_T) \\ \vdots \qquad\qquad \vdots \\ g_m(X_1,\tilde{\alpha}_T,\tilde{\beta}_T) \cdots\cdots g_m(X_T,\tilde{\alpha}_T,\tilde{\beta}_T) \end{bmatrix} \begin{bmatrix} g_1(X_1,\tilde{\zeta}_T,\tilde{\beta}_T) \cdots\cdots g_m(X_1,\tilde{\zeta}_T,\tilde{\beta}_T) \\ \vdots \qquad\qquad \vdots \\ g_1(X_T,\tilde{\alpha}_T,\tilde{\beta}_T) \cdots\cdots g_m(X_T,\tilde{\alpha}_T,\tilde{\beta}_T) \end{bmatrix} \right)^{-1}.$$

$$\begin{bmatrix} g_1(X_1,\tilde{\alpha}_T,\tilde{\beta}_T)\cdots\cdots g_1(X_T,\tilde{\alpha}_T,\tilde{\beta}_T) \\ \vdots \qquad\qquad\qquad \vdots \\ g_m(X_1,\tilde{\alpha}_T,\tilde{\beta}_T)\cdots\cdots g_m(X_T,\tilde{\alpha}_T,\tilde{\beta}_T) \end{bmatrix} \begin{bmatrix} (Y_1-f(X_1,\tilde{\beta}_T)\tilde{\alpha}_T)^2 \\ \vdots \\ (Y_T-f(X_T,\tilde{\beta}_T)\tilde{\alpha}_T)^2 \end{bmatrix}$$

To simplify the notations, let

$$G_T = \begin{bmatrix} g_1(X_1,\alpha_T,\beta_T)\cdots\cdots g_m(X_1,\alpha_T,\beta_T) \\ \vdots \qquad\qquad\qquad \vdots \\ g_1(X_T,\alpha_T,\beta_T)\cdots\cdots g_m(X_T,\alpha_T,\beta_T) \end{bmatrix}$$

and G_T^o, $\tilde{G}_T$ denote G_T evaluated at (α_o,β_o) and $(\tilde{\alpha}_T,\tilde{\beta}_T)$ respectively.

<u>Proposition 2</u>. In the above nonlinear regression system, the two stage estimator $\hat{\gamma}_T$ is strongly consistent if the following conditions are satisfied:

1). $\{\varepsilon_t\}$ and $\{U_t\}$ are independent for different t, zero means and their four mements are uniformly bounded.

2). $f(X_t,\beta)$ and $g(X_t,\alpha,\beta)$ are continuous on $X \times H_1 \times H_2$ where H_1, H_2 are compact subsets and (α_o,β_o) is an interior point in $H_1 \times H_2$.

3). $\{X_t\}$ is bounded and the empirical distribution $G_n(x)$ converges to a distribution.

4). $\lim_{T\to\infty} \frac{1}{T} G_T^{o\prime} G_T^o$ exists and it is a positive definite matrix.

5). $\tilde{\alpha}_T$, $\tilde{\beta}_T$ are strong consistent estimators of α_o and β_o.

Proof: The proof is similar to the proof in proposition 2. It is enough to note in this case,

$$\hat{\gamma}_T = (\tilde{G}_T'\tilde{G}_T)^{-1}\tilde{G}_T'G_T^{o}\gamma + (\tilde{G}_T'\tilde{G}_T)^{-1}\tilde{G}_T'\begin{bmatrix} U_1 \\ \vdots \\ \hat{U}_T \end{bmatrix} + (\tilde{G}_T'\tilde{G}_T)^{-1}\tilde{G}_T\begin{bmatrix} \hat{\varepsilon}_1^2 - \varepsilon_1^2 \\ \vdots \\ \hat{\varepsilon}_T^2 - \varepsilon_T^2 \end{bmatrix}$$

where $\hat{\varepsilon}_T^{\,2} = (Y_t - f(X_t,\tilde{\beta}_T)\,\tilde{\alpha}_T)^2$. Also we have

$$(\tilde{G}_T'\tilde{G}_T)^{-1}\tilde{G}_T\begin{bmatrix} \hat{\varepsilon}_1^2 - \varepsilon_1^2 \\ \vdots \\ \hat{\varepsilon}_T^2 - \varepsilon_T^2 \end{bmatrix} = (\tilde{G}_T'\tilde{G}_T)^{-1}\tilde{G}_T\begin{bmatrix} (f(X_1,\tilde{\beta}_T)\tilde{\alpha}_T)^2 - (f(X_1,\beta_0)\alpha_o)^2 \\ \vdots \\ (f(X_T,\tilde{\beta}_T)\tilde{\alpha}_T)^2 - (f(X_T,\beta_o)\alpha_o)^2 \end{bmatrix}$$

$$+ 2(\tilde{G}_T'\tilde{G}_T)^{-1}\tilde{G}_T\begin{bmatrix} (f(X_1,\beta_o)\alpha_o - f(X_1,\tilde{\beta}_T)\tilde{\alpha}_T)f(X_1,\beta_o)\alpha_o \\ \vdots \\ (f(X_T,\beta_o)\alpha - f(X_T,\tilde{\beta}_T)\tilde{\alpha})f(X_T,\beta_o)\alpha_o \end{bmatrix}$$

$$+ 2(\tilde{G}_T'\tilde{G}_T)^{-1}\tilde{G}_T\begin{bmatrix} (f(X_1,\beta_o)\alpha - f(X_1,\tilde{\beta}_T,\tilde{\alpha}_T))\varepsilon_1 \\ \vdots \\ (f(X_T,\beta_o)\alpha_o - f(X_T,\tilde{\beta}_T)\tilde{\alpha}_T)\varepsilon_T \end{bmatrix}$$

Q.E.D.

Since the two stage estimators are derived under the normality disturbances and under the assumption that all the parameters lie in a compact space, the first four moments of ε_t and U_t are always uniformly bounded.

<u>References</u>

1. Amemiya, T., "Regression Analysis When the Dependent Variable is Truncated Normal," <u>Econometrica</u> 1973, pp. 997-1016.

2. Amemiya, T., "Multivariate Regression and Simultaneous Equation Models When the Dependent Variables are Truncated Normal," <u>Econometrica</u> <u>42</u>, 1974, pp. 999-1012.

3. Amemiya, T., "A Note on a Fair and Jaffee Model," <u>Econometrica</u> <u>42</u>, pp. 759-762.

4. Carliner, G., "Income Elasticity of Housing Demand," <u>The Review of Economics and Statistics</u>, LV, 1973.

5. Fair, R. C. and D. M. Jaffee, "Methods of Estimation for Markets in Disequilibrium," <u>Econometrica</u> 1972, pp. 497-514

6. Fenton, C., "The Permanent Income Hypothesis, Source of Income and the Demand for Rental Housing," in Joint Center for Urban Studies, <u>Analysis of Selected Census Welfare Program Data to Determine Relation of Household Characteristics, Housing Market Characteristics, and Administrative Welfare Policies to a Direct Housing Assistance Program</u> (Draft - Final Report, mimeographed, July 31, 1974).

7. Goldfeld, S. M. and R. E. Quandt, <u>Nonlinear Methods in Econometrics</u> North-Holland Publishing Company, Amsterdam - London, 1972.

8. Goldfeld, S. M. and R. E. Quandt, "Estimation in a Disequilibrium Model and the Value of Information," <u>Journal of Econometrics</u> <u>3</u>, 1975, pp. 325-348.

9. Goldfeld, S. M., H. Kelejiam, and R. E. Quandt, "Least Squares and Maximum Likelihood Estimation of Switching Regressions," Econometric Research Program, Princeton, Research Memorandum No. 130, 1971.

10. Heckman, J., "Shadow Prices, Market Wages and Labor Supply," <u>Econometrica</u>, Vol. <u>42</u> (No. 4), July 1974, pp. 679-694.

11. Heckman, J., "Shadow Prices, Market Wages and Labor Supply Revisited: Some Computational and Conceptual Simplifications and Revised Estimates," June 1975 (manuscript).

12. Jennrich, H. O., "Asymptotic Properties of Non-Linear Least Squares Estimators," <u>Annals of Mathematical Statistics</u> <u>40</u>, (1969), pp. 633-643.

13. Lee, T. H., "Housing and Permanent Income: Tests Based on a Three-Year Reinterview Survey," The Review of Economics and Statistics, L, 1968.

14. Maddala, G. S. and F. D. Nelson, "Maximum Likelihood Methods for the Estimation of Models of Markets in Disequilibrium," Econometrica, November 1974, pp. 1013-1030.

15. McCall, J. J. and S. A. Lippman, "The Economics of Job Search: A Survey," Economic Inquiry, 1976, pp. 155-189 and 347-368.

16. Nelson, F. D., "Estimation of Economic Relationships with Censored, Truncated and Limited Dependent Variables," unpublished Ph.D. Thesis, University of Rochester, 1975.

17. Ohls, J., "A Cross Section Study of Demand Functions for Housing and Policy Implications of the Results," University of Pennsylvania Ph.D. Dissertation, 1971.

18. Poirier, D., "The Determinants of Home Buying in the New Jersey Graduated Work Incentive Experiment," Faculty Working Paper, University of Illinois at Urbana-Champaign, December 1974.

19. Polinsky, A., "The Demand for Housing: A Study in Specification and Grouping," Discussion Paper Number 432, Harvard Institute of Economic Research, Harvard University, Cambridge, Massachusetts, September 1975.

20. Quigley, J., "Housing Demand in the Short Run: An Analysis of Polytomous Choice," Explorations in Economic Research, pp. 76-102, Vol. 3, No. 1, 1976.

21. Reid, M., Housing and Income, Chicago, University of Chicago Press, 1962.

22. Survey Research Center, A Panel Study of Income Dynamics, Institute for Social Research, University of Michigan, 1972.

23. Tobin, J., "Estimation of Relationships for Limited Dependent Variables," Econometrica, 1958.