ON CLASSIFICATION PROBABILITIES FOR THE DISEQUILIBRIUM MODEL

Mark Gersovitz*

Econometric Research Program
Research Memorandum No. 238

December 1978

* I am especially grateful to Richard E. Quandt for his discussion of this paper and many related topics. I would also like to thank Gregory C. Chow, Jonathan Eaton and Stephen M. Goldfeld for their comments.

Econometric Research Program
Princeton University
207 Dickinson Hall
Princeton, New Jersey

ON CLASSIFICATION PROBABILITIES FOR THE DISEQUILIBRIUM MODEL

Mark Gersovitz

In econometric models of markets in disequilibrium, the dependent variable is generated by one of two regimes. Considerable attention has been given to the estimation of these models (Maddala and Nelson, 1974 and Quandt, 1977). Little attention has, however, been directed to the classification of observations into the two different regimes. Yet, this classification is often of considerable interest and can help to provide additional evidence on the plausibility of the model when additional information on the membership of different observations in the different regimes is available. For instance, in Eaton and Gersovitz (1978), the disequilibrium model was applied to credit rationing in international lending to poor countries, and individual countries were then classified as constrained or unconstrained.

This paper discusses two possible methods for classifying observations, investigates the relationship between these two procedures and presents a Monte Carlo evaluation of one of these probability measures. A by-product of this last part of the paper is some evidence on the small sample behavior of the estimators of the model.

1. Classification in the Disequilibrium Model

The simplest disequilibrium model is:

(1) $y_{1t} = \beta_1' X_{1t} + U_{1t}$

(2) $y_{2t} = \beta_2' X_{2t} + U_{2t}$

(3) $y_t = \min(y_{1t}, y_{2t})$

Where β_1 and β_2 are $k_1 \times 1$ and $k_2 \times 1$ coefficient vectors respectively and

X_{1t} and X_{2t} are $k_1 \times 1$ and $k_2 \times 1$ vectors of independent variables. For convenience, I assume that the normal errors, U_{1t} and U_{2t}, are independently distributed with variances σ_1^2 and σ_2^2.

The crucial aspect of this model is that only y is observed, y_1 and y_2 being unobserved. Define

$$(4) \quad f_i(y_t) = \frac{1}{\sqrt{2\pi}\,\sigma_i} \exp\left[\frac{-1}{2\sigma_i^2}(y_t - \beta_i' X_{it})^2\right] \qquad (i=1,2)$$

$$(5) \quad F_i(y_t) = \int_{y_t}^{\infty} f_i(y_{it})\,dy_{it} \qquad (i=1,2)$$

It can be shown that the likelihood of y_t is

$$(6) \quad h(y_t) = f_1(y_t)\,F_2(y_t) + f_2(y_t)\,F_1(y_t)$$

(see Maddala and Nelson, 1974). Maximization of equation (6) yields estimates of the parameters.

It is now possible to ask how an actual observation (y_t, X_{1t}, X_{2t}) can be classified into the two regimes. In this context, two measures are of interest. First, there is the probability that any observation with the given independent variables X_{1t}, X_{2t} will come from regime 1:

$$(7) \quad P(1|X_{1t}, X_{2t}) = \int_{-\infty}^{(\beta_2' X_{2t} - \beta_1' X_{1t})/\sigma} \frac{1}{\sqrt{2\pi}} \exp(-v^2/2)\,dv$$

where $\sigma^2 = \sigma_1^2 + \sigma_2^2$, a formula derived by Maddala and Nelson (1974, p. 1014).

Second, there is the likelihood that the particular observation (y_t, X_{1t}, X_{2t}) did come from regime 1:

$$(8) \quad P(1|y_t, X_{1t}, X_{2t}) = f_1\,F_2/(f_1\,F_2 + f_2\,F_1)$$

where the f_i's and F_i's are defined in equations (4) and (5). This result is easily proved by the manipulation of formulae for conditional probabilities, and has the intuitive rationale of the fraction of the likelihood of the particular observation which was contributed by the first regime.

In general, these probabilities are different since equation (7) uses the realization of the actual y_t. There are a number of results which can be proved about the relationship between these two probabilities.

Define

$$(9) \quad \mu_i = \beta_i' \, x_{it} \qquad (i=1,2)$$

It is obvious from (1) that

$$P(1|x_{1t}, x_{2t}) > .5 \text{ iff } \mu_2 > \mu_1$$

It is then of interest to show that a parallel result holds for $P(1|y_t, x_{1t}, x_{2t})$, but only under very restrictive assumptions.

<u>Theorem</u>: If $\mu_2 > \mu_1$, and $\sigma_1^2 = \sigma_2^2$

Then $P(1|y_t, x_{1t}, x_{2t}) > .5$

<u>Proof</u>: $P(1|y_t, x_{1t}, x_{2t}) > .5$ iff $f_1 F_2 - f_2 F_1 > 0$.

Using the definitions of equations (4) and (5) this last condition holds iff

$$0 > \int_y^\infty \exp \left[\frac{-1}{2} \left\{\frac{(t-\mu_1)^2}{\sigma_1^2} - \frac{(y-\mu_2)^2}{\sigma_2^2} \right\}\right] - \exp \left[\frac{-1}{2} \left\{\frac{(t-\mu_2)^2}{\sigma_2^2} - \frac{(y-\mu_1)^2}{\sigma_1^2} \right\}\right] dt$$

A sufficient condition for the above is

$$\frac{(t-\mu_1)^2}{\sigma_1^2} - \frac{(y-\mu_2)^2}{\sigma_2^2} - \frac{(t-\mu_2)^2}{\sigma_2^2} + \frac{(y-\mu_1)^2}{\sigma_1^2} > 0 \qquad \forall \; t > y$$

Expanding and cancelling terms yields

$$(\sigma_2^2 - \sigma_1^2)\ (t^2 - y^2) - 2(t-y)\ (\sigma_2^2\ \mu_1 - \sigma_1^2\ \mu_2)$$

which is positive for $\sigma_2^2 = \sigma_1^2$ and $\mu_1 < \mu_2$.

Consider the classification criterion that an observation is regarded as coming from regime 1 if its probability of regime 1 exceeds one half. It would not make any difference whether the conditional or unconditional probabilities were used so long as $\sigma_1^2 = \sigma_2^2$. However, if $\sigma_1^2 > \sigma_2^2$ or $\sigma_1^2 < \sigma_2^2$, cases can occur where the two probabilities would imply different classifications. Both situations can be discussed with the aid of Figure 1.

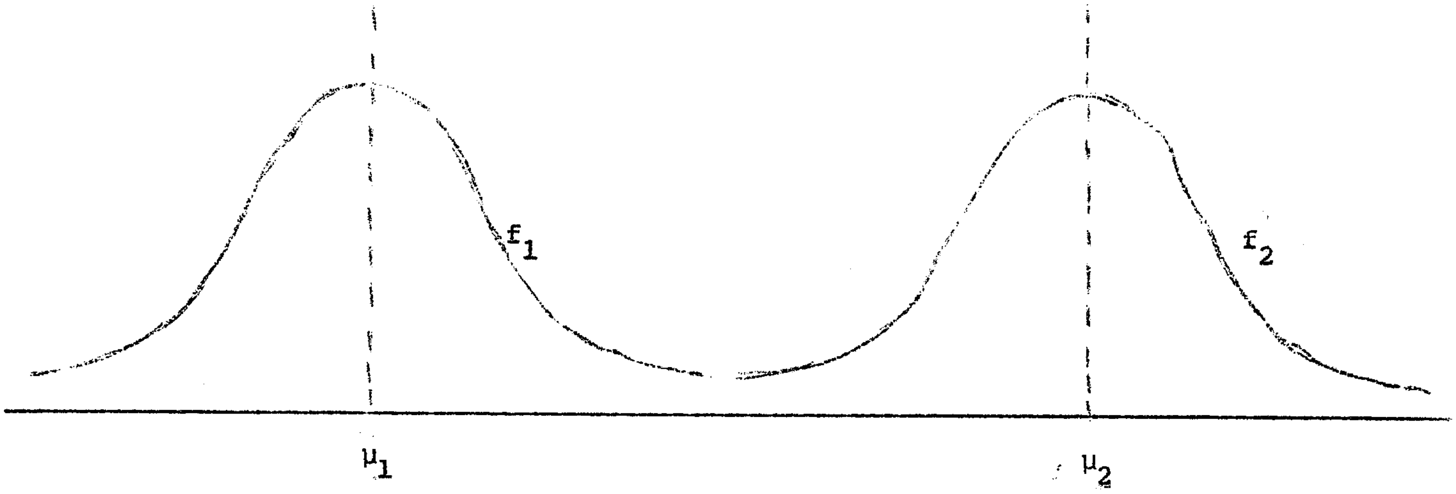

FIGURE 1

For the case $\sigma_1^2 > \sigma_2^2$, $P(1|y_t, X_{1t}, X_{2t})$ exceeds .5 to the left of μ_1 since $f_1 > f_2$ and $F_1 < F_2$. Between μ_1 and μ_2, this probability declines since $dP/dy_t = f_1' f_2 - f_2' f_1$ which is negative. The rate of decline can be so steep that $P(1|y_t, X_{1t}, X_{2t}) < .5$ for $y_t < \mu_2$. One set of values which produces this effect is: $\mu_1 = 0$, $\mu_2 = 1$, $\sigma_1 = 1.5$, $\sigma_2 = .25$.

On the other hand, for $\sigma_2^2 > \sigma_1^2$, $P(1|y_t, x_{1t}, x_{2t})$ can be less than .5 to the left of μ_1 where it is now possible that $f_2 > f_1$ and $F_2 < F_1$. One set of values which produces this effect is: $\mu_1 = 0$, $\mu_2 = 1$, $\sigma_1 = .25$ and $\sigma_2 = 1.75$.

These examples point out the fact that these two probability measures will only yield the same classification under very special circumstances. Each measure is valid in response to its own question -- conditional or unconditional classification. It is the purpose of the preceding discussion to emphasize the need to enunciate this question explicitly before choosing one or the other measure. In the next section I focus on the conditional probability measure because it has not been previously discussed as a classification criterion and because I believe the question it answers is of considerable interest in many applied contexts.

2. Monte Carlo Studies of the Conditional Classification Probability

Because $P(1|y_t, x_{1t}, x_{2t})$ must be evaluated using estimated values of the parameters, I investigated the behavior of the estimator of $P(1|y_t, x_{1t}, x_{2t})$. So long as estimates of the model are derived via maximum likelihood, the estimates of this probability using the estimated parameters will be maximum likelihood. This property follows because functions of maximum likelihood estimators are maximum likelihood estimators of the function. The estimates of $P(1|y_t)$ have the usual maximum likelihood properties. (From this point, I suppress the arguments x_{1t}, x_{2t}.)

The purpose of this section is to go beyond the maximum likelihood characterization by investigating the small sample properties of these probability estimators using Monte Carlo methods. In particular, two hypotheses are of special interest. First is that there may be systematic bias in the estimates of the probability dependent on the true value of the probability. Low probabilities may be biased toward zero and high probabilities toward one. Hartigan (1975, p. 120) in investigating classification in a different but related model stated: "Usually, the probabilities that each case belongs to various clusters are very close to unity or zero, but this should not be taken as an indication of sharply defined clusters." While Hartigan does not develop this observation, the phenomenon clearly deserves investigation.

A second hypothesis is that while the probability estimators may be biased or unbiased, they may have greater variance for values in the middle range. This hypothesis is analogous to the result for the estimation of a simple population probability π the variance of which $\pi(1-\pi)/n$, is at a maximum for $\pi = .5$.

The form of the model used in the Monte Carlo experiments is

$$y_{1t} = \beta_{11} + \beta_{12} x_{1t} + u_{1t} \tag{10a}$$

$$y_{2t} = \beta_{21} + \beta_{22} x_{2t} u_{2t} \tag{10b}$$

$$y_t = \min (y_{1t}, y_{2t}) \tag{10c}$$

$$\sigma_2 = k \sigma_1 \tag{10d}$$

The additional constraint (10d) where k is a fixed constant known to the investigator, is added to ensure that the likelihood is bounded (see Quandt, 1977 for a discussion of this problem). Consequently, all replications are calculated optimizing (6) subject to (10d) where k is the true ratio of σ_2 to σ_1. An alternative approach is to maximize the likelihood by searching for a local maximum without imposing (10d). Amemiya and Sen (1977) show that there exists a local maximum which yields consistent parameter estimates. In an actual estimation situation where estimation need only be done once, this method is feasible despite the care necessary to avoid initial conditions leading to the global maximum. In a Monte Carlo situation, however, the number of replications prohibits consideration of each individual maximization history to ensure that the algorithm does not enter the unbounded region. To maintain computational feasibility, I therefore imposed (10d).

A particular version of (10) requires a choice for the four β_{ij}'s, σ_1, k and the X_{1t}, X_{2t}'s. The X_{1t}'s and X_{2t}'s were generated by a normal distribution and were kept fixed in all replications except for t=1 as discussed below. A replication encompasses a random drawing of the U_{1t}'s and U_{2t}'s for a sample size of T and the estimation of β_{ij}'s and σ_1 to yield observations on the small sample estimators.

The errors, U_{1t} and U_{2t} must be constant if the true $P(1|y_t)$ is to be the same for each of the N replications. Consequently, I fixed U_{1t}, U_{2t} for t=1, allowing the other U_{1t} and U_{2t} to vary randomly over N replications and then observed the statistics on $P(1|y_t, t=1)$ over the set of replications. To generate statistics on a different value of $P(1|y_t, t=1)$ a new set of observations on X_{1t}, X_{2t}, U_{1t}, U_{2t} for t=1 was drawn and the N replications were repeated using the same values for the X_{1t}, X_{2t}, U_{1t} and U_{2t} for t>1. Finally, the whole process was repeated for a different value of k using the same values for the X_{1t}, X_{2t}, U_{1t} and U_{2t} t=1 ... T. Prior to the examination of the small sample properties of the estimator of $P(1|y_t)$, I allowed the U_{1t} and U_{2t} for t=1 to vary randomly at each replication to investigate the small sample behavior of the estimators of the β_{ij}'s and σ_1. (These results are presented in Tables 1 and 2.)

The small sample properties of the estimators were examined with antithetic variates. This method has the potential for considerable reduction in the number of replications necessary to achieve a given variance of the mean of the estimator (see Hendry and Harrison, 1974). To implement the antithetic variates approach, consider the observation on the estimator $\hat{\theta}$ of a parameter obtained on the ith replication, denoted θ_i. Replace the error terms U_{1t} and U_{2t}

TABLE 1

Coefficient Biases: Sample Size = 30

Experiment 1: $\sigma_2 = \sigma_1$

Coefficient	True Value	Bias	t-Statistic
β_{11}	0	.032	1.38
β_{12}	1	.014	1.66
β_{21}	0	.102	4.29
β_{22}	-1	-.030	4.46
σ_1	1	.096	5.25
σ_2	1		

Experiment 2: $\sigma_2 = .5\sigma_1$

Coefficient	True Value	Bias	t-Statistic
β_{11}	0	.073	2.68
β_{12}	1	.026	2.77
β_{21}	0	.030	1.86
β_{22}	-1	-.009	1.91
σ_1	1	.049	2.41
σ_2	.5		

TABLE 2

Coefficient Biases: Sample Size = 90

Experiment 1: $\sigma_2 = \sigma_1$

Coefficient	True Value	Bias	t-Statistic
β_{11}	0	.036	3.13
β_{12}	1	.015	4.13
β_{21}	0	-.005	.35
β_{22}	-1	-.001	.26
σ_1	1	-.031	2.92
σ_2	1		

Experiment 2: $\sigma_2 = .5\sigma_1$

Coefficient	True Value	Bias	t-Statistic
β_{11}	0	.011	.99
β_{12}	1	.005	1.31
β_{21}	0	.002	.30
β_{22}	-1	-.001	.58
σ_1	1	.003	.27
σ_2	.5		

in (10) by their negatives and repeat the replication yielding an observation θ'_i on the expected value of estimator $\hat{\theta}$. For the problem of (10) $E(\theta_i) = E(\theta'_i) = E(\hat{\theta})$. Now take as the observation on $\hat{\theta}$ from the ith pair of replications.

$$(11) \qquad \overline{\theta}_i = (\theta_i + \theta'_i)/2$$

Clearly $E(\overline{\theta}_i) = E(\hat{\theta})$. Further if the θ_i and θ'_i are sufficiently negatively correlated, the variance of $\overline{\theta}_i$ will be much less than the variance of θ_i and considerable gains in efficiency will be realized.

For the experiments which follow, the errors were generated using a multiplicative congruential method to generate uniformly distributed errors. These uniform variates were converted to normal variates using the method of Box and Muller (1958). All calculations were done in double precision FORTRAN using the IBM 360/91 at Princeton University. Estimation of the parameters for each replication was done using the GRADX option of the GQOPT program based on the method of Goldfeld, Quandt and Trotter (1966). The number of replications was N=50 and the sample size for each experiment was either T=30 or T=90.

The true values of the β_{ij} used in generating the 50 replications on each of the four variants of (10) are shown in Table 1. The X_{it} were generated by drawing from a normal distribution with mean zero and standard deviation 2.5 and were held constant across replications and experiments. The X_{1t} and X_{2t} were modified to have a correlation coefficient of $\sqrt{.5}$. The U_{1t} were generated by a standard normal. The U_{2t} were normal deviates independent of the U_{1t} with standard deviation equal to or half that of the standard deviation of U_{1t} as indicated in the Tables. Consequently, the implicit, theoretical $R^2[= 1-\sigma_u^2/(\sigma_u^2+\sigma_x^2)]$ of each equation is between .85 and .95.

Tables 1 and 2 give the bias (the difference between the average $\overline{\theta}_i$, the antithetic estimate and the true value) for the experiments on the β_{ij}. The t-statistic tests whether the bias is significantly different from zero using the formula for the standard deviation of a mean, $\bar{\sigma}/\sqrt{N}$ where $\bar{\sigma}$ is the standard deviation of the $\bar{\theta}_i$'s and $N=50$. The ratio to the standard deviation of the θ_i, the non-antithetic estimates (not shown), to the $\bar{\sigma}$ was between 2 and 5 to 1. Consequently, the precision of the antithetic variates is equivalent to a non-antithetic study using 200 to 1,250 replications. (This equivalence is calculated using the formula for the standard deviation of the mean of the estimates, $\sigma/\sqrt{N}$, and solving for the N required to offset the larger, non-antithetic σ.) The biases exhibited in Table 1 are largely significant but are not severe in magnitude. These biases fall in both significance and magnitude as the sample size is tripled (see Table 2).

Note that the bias for the regime 2 coefficients falls from experiment 1 to 2 in each table as σ_2 falls. The biases for the regime 2 coefficients in Table 1 are more severe than the biases for the regime 1 coefficients. Although the true model is symmetric with respect to the two regimes, the particular X_{it}'s which are kept constant from replication to replication, can lead to this asymmetry with respect to relative biases. In practice, roughly half the y_t's belong to each regime, however.

Tables 3 and 4 give the results for $P(1|y_t)$ estimates corresponding to Tables 1 and 2. The first column gives the true $P(1|y_t, t=1)$ [i.e. calculated using the true parameters of the model in equation (8)]. This value was held constant for N replications as described previously. The next column gives the average of the N antithetic observations on the estimator of the probability [i.e. calculated using the estimated parameters of the model in equation (8)]. The bias records the difference between columns 1 and 2.

The standard deviation of the antithetic estimates of $P(1|y_t)$ is given by $\bar{\sigma}$, and $\sigma_{\hat{p}}$ gives the standard deviation of the θ_i estimates. The $\sigma_{\hat{p}}$'s are relevant to an assessment of the variability of the estimator of $P(1|y_t)$

TABLE 3

Probability Estimators: Sample Size = 30

Experiment 1: $\sigma_1 = \sigma_2$

True Probability	Antithetic Estimated Probability	Bias	t-Statistic	$\bar{\sigma}$	$\sigma_{\hat{p}}$
.0025	.0057	.0032	2.76	.0082	.0117
.085	.114	.029	3.27	.063	.130
.505	.532	.027	2.72	.070	.285
.540	.567	.027	1.97	.097	.365
.754	.782	.028	2.36	.084	.189
.869	.818	-.051	3.68	.098	.195
.959	.954	-.005	1.16	.030	.065

Experiment 2: $\sigma_1 = .5\sigma_2$

True Probability	Antithetic Estimated Probability	Bias	t-Statistic	$\bar{\sigma}$	$\sigma_{\hat{p}}$
.00053	.00089	.00036	1.59	.0016	.0015
.045	.049	.004	.78	.033	.059
.575	.548	-.029	1.60	.128	.320
.744	.717	-.027	1.48	.129	.254
.826	.839	.013	1.08	.085	.139
.994	.968	-.026	3.91	.047	.083
.9998	.9989	-.0009	2.36	.0027	.0044

TABLE 4

Probability Estimators: Sample Size = 90

Experiment 1: $\sigma_1 = \sigma_2$

True Probability	Antithetic Estimated Probability	Bias	t-Statistic	$\bar{\sigma}$	$\sigma_{\hat{p}}$
.0025	.0027	-.0002	.61	.0023	.0038
.085	.082	-.003	.82	.026	.059
.505	.497	-.008	1.30	.044	.149
.540	.522	-.018	1.90	.067	.226
.754	.751	-.003	.45	.047	.115
.869	.834	-.035	4.58	.054	.127
.959	.952	-.007	2.25	.022	.039

Experiment 2: $\sigma_1 = .5\sigma_2$

True Probability	Antithetic Estimated Probability	Bias	t-Statistic	$\bar{\sigma}$	$\sigma_{\hat{p}}$
.00053	.00059	.00006	.70	.00061	.00085
.045	.044	-.001	.60	.013	.024
.575	.571	-.004	.35	.071	.162
.744	.741	-.003	.27	.073	.125
.826	.830	.004	.52	.053	.081
.994	.983	-.011	3.12	.025	.023
.9998	.9994	.0002	.80	.0018	.0011

since in an actual estimation situation a θ_i rather than a $\bar{\theta}_i$ estimate is formed. Again, the ratio of $\bar{\sigma}$ and $\sigma_{\hat{p}}$ indicates that, for most cases, the gains in the precision of the estimate of the bias derived from the antithetic approach are quite considerable.

The t-statistic gives the bias divided by its standard deviation $\bar{\sigma}/\sqrt{N}$. The bias is considerably more pronounced for the small sample experiments of Table 1, but neither table exhibits extreme bias. There is some tendency in the small sample (T=30) case for low values of the probability to be overestimated and high values to be underestimated. This phenomenon contradicts the supposition, mentioned at the beginning of this section, that the probability estimates might give a spuriously sharp separation of the sample by drifting toward zero or one. The bias is generally larger in absolute magnitude for the intermediate values of the true $P(1|y_t)$.

The variability of the estimator, $\sigma_{\hat{p}}$, is also large for intermediate values of the true $P(1|y_t)$. This result substantiates the heuristic argument made above by analogy with the variance of the estimate of a population probability. For these intermediate values of $P(1|y_t)$, $\sigma_{\hat{p}}$ can be disturbingly large, suggesting rather erratic behavior.

3. Conclusions

This paper has distinguished between conditional and unconditional methods of classifying observations in models where membership in either of two regimes is determined by a minimum condition. The relation of the two measures to each other is discussed. In general, there is no necessary correspondence between the two methods so that care must be taken in choosing the relevant one in any application.

I then focused on the conditional classification probability because it is important in many applied contexts to assign the observations used in estimation between the two regimes. Monte Carlo results indicated that

1. bias need not be serious in even relatively small samples and diminishes rapidly as sample size increases. In small samples bias tends to be largest for intermediate values of the true probability.

2, low values of the true probability tend to be overestimated while high values are underestimated so that the partition of the sample does not tend to be spuriously sharp.

3. the variability of the estimated probability is largest for intermediate values of the true probability.

Consequently it appears that the estimated probabilities are most reliable, both in terms of bias and variability, when they take extreme rather than intermediate values.

References

Amemiya, T. and G. Sen (1977). "The Consistency of the Maximum Likelihood Estimator in a Disequilibrium Model," Institute for Mathematical Research in the Social Sciences, Technical Report No. 238, Stanford University.

Box, George E. P. and Mervin E. Muller (1958). "A Note on the Generation of Normal Deviates," Annals of Mathematical Statistics, 29, pp. 610-1.

Eaton, Jonathan W. and Mark Gersovitz (1978). "Debt with Potential Repudiation: Theoretical and Empirical Analysis," unpublished mimeo.

Goldfeld, Stephen M., Richard E. Quandt and Hale F. Trotter (1966). "Maximization by Quadratic Hill Climbing," Econometrica, 34, pp. 541-551.

Hartigan, John A. (1975). Clustering Algorithms, New York: Wiley.

Hendry, David F. and Robin W. Harrison (1974). "Monte Carlo Methodology and the Small Sample Behaviour of Ordinary and Two-Stage Least Squares," Journal of Econometrics, 2, pp. 151-174.

Maddala, G. S. and Forrest D. Nelson (1974). "Maximum Likelihood Methods for Models of Markets in Disequilibrium," Econometrica, 42, pp. 1013-1030.

Quandt, Richard E. (1977). "Maximum Likelihood Estimation of Disequilibrium Models," Econometric Research Program Paper, Princeton University.