COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 803

TEMPORAL DEPENDENCE IN LIMITED DEPENDENT VARIABLE MODELS: THEORETICAL AND MONTE-CARLO RESULTS

Vassilis Argyrou Hajivassiliou

August 1986

Temporal Dependence in Limited Dependent Variable Models:

Theoretical and Monte-Carlo Results

by

Vassilis Argyrou Hajivassiliou*

Department of Economics

Yale University

October 1985

Revised August 1986

ABSTRACT

This paper analyzes the consistency properties of classical estimators for limited dependent variables models, under conditions of serial correlation in the unobservables. A unified method of proof is used to show that for certain cases (e.g., Probit, Tobit and Normal Switching Regimes models, which are normality-based) estimators that neglect particular types of serial dependence (specifically, corresponding to the class of "mixing" processes) are still consistent. The same line of proof fails for the analogues to the above models that impose logistic distributional assumptions, thus indicating that normality plays a special role in these problems. Sets of Monte-Carlo experiments are then carried out to investigate these theoretical results.

*This paper draws on results from Chapter 2 of my doctoral dissertation, and was presented at the 1985 Winter meetings of the Econometric Society. I would like to thank Don Andrews, Jerry Hausman, Dan McFadden, Jim Powell, Whitney Newey and Hal White for very helpful comments and suggestions. Any remaining errors and shortcomings are, of course, my own.

Temporal Dependence in Limited Dependent Variable Models: Theoretical and Monte-Carlo Results

Introduction:

Consistency under "Mixing" Conditions

Analysis of limited dependent variable models in the presence of serial correlation in the unobservables reveals that full ML estimation is in general very intractable, because multivariate integration (in some cases T-fold) is involved in evaluating the likelihood function.

Investigators in this area have mainly taken two routes in meeting this problem: first, attempts have been made to present computationally feasible estimators that are appropriate for such models, though they are statistically inefficient relative to MLE. See Avery, Hansen and Hotz (1983) and Ruud (1981) for orthogonality-conditions (OC) type of estimators.

Second, the behaviour of the ordinary (misspecified) ML estimator that wrongly treats the errors as serially independent has been studied (Robinson (1982), Avery, Hansen and Hotz (1983)). The main conclusion from this line of research is that for certain limited dependent variable models (in particular for Probit, Tobit and normal Switching), the misspecified MLE is still consistent and asymptotically normal, with an asymptotic covariance matrix that differs from the one obtained under the incorrect assumption of i.i.d. errors.

This conclusion of the latter approach could lead to a premature belief that failure to treat serial correlation correctly may not be too serious after all, since consistency of the estimator is preserved and hence correct inference only requires consistent estimates for the standard errors of the standard estimator.[1] In this paper we show that, rather surprisingly, this consistency result does not in general extend to other widely used estimators for LDV models that do not work within the normal family of distributions. We illustrate this for the cases of the binomial Logit model (McFadden (1973)) and the analogues to the Tobit and switching models that assume logistically distributed errors (which we call L-tobit and L-switching respectively).

In Section I we adapt results in Levine (1983), White and Domowitz (1984), Newey and Powell (1985), and White (1985) to prove the consistency of the estimators that rely on normality and to highlight the requirements for consistency. Compared to the proof given by Robinson (1982) for the Tobit model, the proof here is much more straightforward, and readily extends to other limited dependent variables models in the presence of serial correlation. We illustrate this proof for the Probit (Finney (1964)) and Tobit models with autoregressive errors. Moreover, we discuss the factor-analytic error structure appropriate for longitudinal data analysis.

We proceed in Section II to analyze along the same lines the cases of Logit and L-tobit. It is shown that the consistency requirements then fail to be satisfied under ARMA-type logistic error specifications, since under such conditions the regression function and the conditional density for a given time period do not in general have the logistic form. Hence, one conclusion that can be drawn is that the generally acknowledged closeness of the Probit and Logit estimates (the two approximately differing by a scale factor of $\pi/\sqrt{3}$ that standardizes for the unidentified variances-see Amemiya (1981)) may not hold in the presence of serial correlation of the ARMA type.

Since what we show here is the failing of sufficient conditions for a proof of consistency under logistic distributional assumptions, we proceed to examine the <u>operational characteristics</u> of logit-like estimators under time-dependence conditions, as compared to their normality-based counterparts. In Section III we present such a comparison of logit and probit estimators under serial dependence using sets of Monte-Carlo experiments.

We conclude with a summary of the results of this paper and with an agenda for future research.

Section I:

Consistency in The Case of Normality Under False i.i.d. Assumptions

There have been several attempts to examine the effects and the possibility of introducing serial dependence into the errors of limited dependent variables models. Quandt (1981) allows for both demand and supply shocks to follow AR(1) processes in the simple disequilibrium model with no observations on regimes (see Fair and Jaffee (1972)); he then proceeds to apply his estimator to a model of the U.S. labour market, claiming to have established that the MLE that incorporates serial correlation in the switching model, although computationally laborious (each function evaluation requiring 2T quadratures), is feasible. As Lee (1984a) points out, however, Quandt's approach was not the true MLE; Lee then presents the correct likelihood expression, and thus shows that the MLE is indeed computationally infeasible because each function evaluation requires two T-fold integrals. Robinson (1982) for the Tobit model, and Avery et al. (1983) for the multiperiod (panel) Probit also recognize the intractability of MLE. They therefore investigate the properties of MLE under (neglected) dependence conditions. Robinson (1982) also conjectures that his consistency proof could be adapted to cover the switching regressions model.

Without concentrating on any specific model in particular, White and Domowitz (1984), Levine (1983) and White (1985) examine the effects of such misspecification with respect to untenable i.i.d. assumptions for the method of Maximum Likelihood Estimation. We here offer a general proof of consistency for a particular class of serially dependent stochastic processes called "mixing" processes, which includes the widely considered class of (finite) ARMA processes, as well as factor-analytic error structures. This method of proof combines theoretical results in Levine (1983), White and Domowitz (1984), White (1985) and Newey and Powell (1985). A novel feature for cases with concave criterion function, is the removal of the usual requirement that the parameter space be compact.

The consistency theorem we present implies that normality assumptions may be of very special importance in establishing consistency properties of the misspecified MLE because they generally yield marginal densities still of the normal form. Specifically there appear to be internal consistency problems under logistic assumptions for ARMA-type processes.

Consider the standard Probit and Tobit models, with y_t , x_t being the observables, and y_t^* being a latent variable:

(1) Probit:
$$y_t^* = x_t' \beta + \varepsilon_t$$
$$y_t = 1, \text{ iff } y_t^* > 0,$$
$$= 0 \text{ otherwise}$$

(2) Tobit:
$$y_t^* = x_t' \beta + \varepsilon_t$$
$$d_t = 1,\ y_t = y_t^* \text{ iff } y_t^* > 0,$$
$$d_t = 0,\ y_t = 0 \text{ otherwise}$$

ε_t is assumed to be i.i.d. $N(0, \sigma^2)$ in both models. Since in the Probit case the scale of y_t cannot be identified, σ^2 is normalized at 1. Under these assumptions, the MLE estimators are defined by:

(3) $$\tilde{\theta} = \underset{\beta}{\operatorname{argmax}} \left\{\frac{1}{T} \sum_{t=1}^{T} \log f_t\right\} ,$$

where

(4) $$f_t = \Phi(x'_t\beta)^{y_t} \cdot (1 - \Phi(x'_t\beta))^{1-y_t} \quad \text{for the Probit,}$$

and

(4') $$f_t = (1 - \Phi(x'_t\beta/\sigma))^{1-d_t} \cdot (\phi((y_t - x'_t\beta)/\sigma)/\sigma)^{d_t} \quad \text{for the Tobit.}$$

$\Phi(.)$ and $\phi(.)$ are the cumulative distribution function and the probability density function for the normal distribution respectively.

We present a theorem that can be applied to show that the ordinary Probit and Tobit estimators are strongly consistent under time dependence conditions. Consider the stochastic process $Z_t = \{(Y_t,X_t)\}$ with Y_t endogenous and X_t predetermined at time t. Suppose $\{Z_t\}$ is stationary so that we can talk about a typical time period t. Let $f(y_t|x_t,\theta_o)$ be the actual density of Y_t conditional on X_t, where θ_o is a p-dimensional column vector. Define the Quasi loglikelihood contribution by

(5) $$\lambda^t(\theta) = \log f(y_t|x_t,\theta)$$

and let

(6) $$\tilde{\theta}_T(\text{QMLE}) = \underset{\theta}{\text{argmax}} \; \{L_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} \lambda^t(\theta)\}$$

Denote the score contributions by $\lambda^t_\theta(\theta)$.

We consider the following set of assumptions:

Assumption (1): *$\{Z_t\}$ is stationary and ϕ-mixing of size $r/(2r-1)$, $r \geq 1$, or α-mixing of size $r/(r-1)$, $r > 1$.*

Definition 1: A stochastic process $\{Z^t\}$ is strong or α-mixing if and only if there exists an infinite sequence of scalars $\alpha = (\alpha_o,\alpha_1,...)$ with $\lim \alpha_m = 0$ such that, for any event F^t defined by $Z^1,...,Z^t$ and event F^{t+k} defined by Z^{t+m}, Z^{t+m+1},...

$$|\text{prob}(F^{t+m},F^t) - \text{prob}(F^{t+m}) \cdot \text{prob}(F^t)| \leq \alpha_m .$$

Definition 2: A uniform or ϕ-mixing stochastic process is characterized by an infinite sequence ϕ_m, specified exactly as sequence α_m in Definition 1, satisfying

$$|\text{prob}(F^{t+m}|F^t) - \text{prob}(F^{t+m})| \leq \phi_m, \ \text{prob}(F^t) \neq 0 .$$

Definition 3: Let the stochastic process $\{Z^t\}$ be uniform ϕ- (strong α-) mixing with the ϕ- (α-) coefficient being an infinite sequence of order $O(m^{-s})$. Then s is termed the size of the mixing process.

Remark 1: m gives the minimum time separation between subsamples of which we try to measure the dependence. To fix ideas, an MA(τ) process would have $\phi(m) > 0$, $\alpha(m) > 0$ for $m < \tau$ and both $\phi(m)$ and $\alpha(m)$ equal to 0 for any $m > \tau$. Such processes are said to be τ-dependent (Billingsley (1979)). The requirement that a process be mixing is akin to a requirement of asymptotic independence. From Definition 3 we see that the mixing size gives the rate at which dependence is vanishing. Hence, with separation of periods equal to m, the mixing coefficients go to 0 at rate $O(m^{-s})$, if the size is s.

Remark 2: ϕ-mixing is a stronger requirement that α-mixing, in that ϕ-mixing implies α-mixing. This is clearly reflected in the size requirements: since ϕ-mixing is stronger than α-mixing, we only need the process to be of size $r/(2r-1)<r/(r-1)$.

Assumption (2): *$f(y_t|x_t, \theta_o)$ is the true conditional density of Y_t given X_t and is a.e. equal to $f(y_t|x_t, \theta)$ only if $\theta = \theta_o$ (global identification).*

Assumption (3): *$\{\lambda^t(\theta)\}$ is a random sequence dominated in absolute value by uniformly integrable functions. Specifically the following domination of moments ($\delta > 0$) is assumed:*

$$\underset{\theta}{E\sup}|\lambda^{t}(\theta)|^{r+\delta} < \Delta < \infty ,$$

where $r > 1$ for strong mixing and $r \geq 1$ for uniform mixing.

<u>Assumption (4)</u>: *θ_o lies in the interior of Θ , a compact subspace of $\mathbb{R}^p$.*

<u>Assumption (5)</u>: *$\lambda_t(\theta)$ is twice continuously differentiable on Θ.*

<u>Assumption (6)</u>: *$L_{\theta\theta}(\theta_o)$ is a pxp negative definite matrix (local identification).*

<u>Theorem 1</u>: If Assumptions (1)-(6) hold, $\tilde{\theta}_T(\text{QMLE}) \xrightarrow{a.s.} \theta_o$.

<u>Proof</u>: Assumptions (1)-(5) are sufficient for the Generic Uniform Law of Large Numbers (Theorem 1, p.3) of Andrews (1986), thus implying that $L_T(\theta) \xrightarrow{a.s.} L(\theta)$ uniformly in θ on Θ , where $L(\theta) \equiv E\lambda^t(\theta)$. By the global and local identification Assumptions (2) and (6), $L(\theta)$ attains a unique maximum at θ_o , and by Assumption (4), Lemma 3 of Amemiya (1973), and the definition of $\tilde{\theta}_T(\text{QMLE}) = \text{argmax } \{L_T(\tilde{\theta})\}$, $\tilde{\theta}_T$ (QMLE) then exists for large enough T and converges to θ_o almost surely.

We now state a Lemma that will allow us to relax the usual compactness requirement of Assumption (4), in cases where the maximand is globally concave in θ (e.g., the Probit model):

<u>Lemma 1</u>: (Newey and Powell (1985)):

Let θ_o be a point in $\mathbb{R}^p$ and Θ be an open set containing θ_o. If

(a) $L_T(\theta)$ converges almost surely to $L(\theta)$ for all θ in $\mathbb{R}^p$ uniformly on Θ,

(b) $L(\theta)$ has a unique maximum on Θ at θ_o.

(c) $L_T(\theta)$ is concave in θ.

Then for $\tilde{\theta} = \text{argmax}_{\theta \in \mathbb{R}^p} L_T(\theta)$,

(i) $\tilde{\theta}$ exists for large enough T, almost surely.

(ii) $\tilde{\theta}$ converges almost surely to θ_o.

Consider the following assumptions:

Assumption (4'): *θ_o lies in the interior of Θ, an open set in $\mathbb{R}^p$.*

Assumption (6'): *$L_{\theta\theta}(\theta)$ is a pxp negative definite matrix (concavity).*

We then obtain:

Theorem 2: If Assumptions (1)-(3), (4'), (5), and (6') hold, $\tilde{\theta}_T(\text{QMLE}) \xrightarrow{a.s.} \theta_o$.

Proof: The Assumptions are sufficient for the Generic Law of Large Numbers of Andrews (1986), and for Lemma 1 of Newey and Powell (1985).

COMMENTS: 1. It is important to note that Assumption (2) requires only that $f(y|x)$ be the correct conditional density, not that $\prod_{t=1}^{T} f(y_t|x_t)$ be the right likelihood for the sample $\{Z_1,\ldots,Z_T\}$. In the presence of serial correlation clearly $\prod_{t=1}^{T} f(y_t|x_t)$ is not the right likelihood since each contribution is not independent from the others. Indeed, $\prod_{t=1}^{T} f(y_t|x_t)$ may not correspond to the correct likelihood for any model -- we only need the conditional density to be specified correctly.

2. Common proofs of consistency for nonlinear models (e.g., White and Domowitz (1984), and Levine (1983)) generally employ a uniform Law of Large Numbers due to Hoadley (1971) which, as pointed out by Andrews (1986), is not applicable when the criterion function is unbounded. Such proofs therefore rule out interesting cases like Quasi-ML models with normal errors.

3. With a strengthening of Assumption (3) to cover the domination of $\{\lambda_\theta^t\}$ and $\{\lambda_{\theta\theta}^t\}$, one could obtain Theorems proving the asymptotic normality of the QML estimators. The proof is a trivial modification of the one used by Levine (1983) in establishing his Theorem 2. We do not give the details here because we are concerned only with consistency properties.

Proof of Consistency for Probit and Tobit with AR(1) and Factor-Analytic Errors:

Proving the consistency of the Tobit (and Probit) estimator(s), which was the main focus of Robinson (1982), becomes a simple exercise of checking the conditions of Theorem 1. As an illustration, consider the AR(1) process $\varepsilon_t = \rho\varepsilon_{t-1} + \eta_t$ with normal innovation $\eta_t \sim N(0, \sigma_\eta^2)$, in Models (1) and (2). We first impose the following structure on the distribution of the exogenous variables:

Assumption (7): *The stochastic process $\{X_t\}$ is stationary and strong (α-) mixing.*

Assumption (8): *$E\, X_t X_t'$ is a positive-definite pxp matrix.*

The proof consists of checking Assumptions (1) to (6):

An AR(1) with normal innovation is a finite order Gaussian ARMA process and is therefore α-mixing with an exponentially decaying coefficient $\alpha(m) \lesssim k \cdot \exp(-\lambda m)$, $\lambda > 0$. See Ibragimov and Linnik (1971). (This implies, in particular, that we can take the parameter r in Assumption (1) arbitrarily close to 1.) Together with Assumption (7) this then ensures that the observable $\{Z_t\}$, being a (measurable) non-linear function of mixing processes is a mixing process itself, by Lemma 2.1 of White and Domowitz (1984). Hence Assumption (1) is satisfied.

By the properties of the Normal distribution:

$y_t^* \sim N(x_t'\beta \, , \, \sigma_\varepsilon^2)$, where $\sigma_\varepsilon^2 = \sigma_\eta^2/(1 - \rho^2)$ by stationarity ($|\rho|<1$).

Hence for the Probit,

$$(7) \quad f(y_t|x_t, \beta, \sigma_\varepsilon^2 = 1) = \Phi(x_t'\beta)^{y_t} \cdot (1 - \Phi(x_t'\beta))^{1-y_t}$$

with normalization $\sigma_\varepsilon^2 = \sigma_\eta^2/(1 - \rho^2) = 1$, and

for the Tobit

$$(7') \quad f(y_t|x_t,\beta,\sigma_\varepsilon^2) = (1 - \Phi(x_t'\beta/\sigma_\varepsilon))^{1-d_t} \cdot (\phi((y_t - x_t'\beta)/\sigma_\varepsilon)/\sigma_\varepsilon)^{d_t} .$$

Note that in both models that parameter ρ is unidentified with respect to the marginal distributions.[2]

Since $EL_T(\theta)$ exists, the global identification requirement can be established, by first noting that it may be translated to:

For $|\theta - \theta_o| \geq \varepsilon > 0$, there exists $\delta > 0$ such that $E(L_T(\theta_o) - L_T(\theta)) \geq \delta$,

where the expectation is taken over the distribution of the vector of observables (Y,X).

For the <u>Probit case</u> the quasi-log likelihood contribution is globally concave, achieving a unique maximum at the true parameter vector. (See Pratt (1981) for an explicit discussion of the concavity properties of limited dependent variable models.)

In the <u>Tobit case</u>, the expressions are exactly the same as in the i.i.d. case, and the reader is referred to Amemiya (1973). Further, in analogy to the Probit case, one can show that $L(\theta)$ reaches a (global) maximum at $\theta = \theta_o$ by establishing the definiteness of the reparameterized Hessian matrix $L_{\theta\theta'}(\theta')$, where $\theta' = (\beta/\sigma, 1/\sigma)$ as in Olsen (1978). Since this reparameterization is a monotonic (separately in both parameters), one-to-one and onto transformation of the original θ, the unimodality (though not the concavity) of the θ parameterization is proved.

In view of the (exponentially declining α-) mixing properties of the Gaussian AR(1) ϵ_t, Assumption (3) is straightforward to establish: Under normality, the error has all moments bounded. Then recall that

$$(8) \quad \lambda^t(\theta) = y_t \log \Phi(x_t'\beta) + (1 - y_t) \log(1 - \Phi(x_t'\beta))$$

We must therefore examine the convergence of terms of the form

$$T^{-1} \sum_{t=1}^{T} \log \Phi(x_t'\beta) \log(1 - \Phi(x_t'\beta)) .$$

This holds under the mixing conditions on $\{Z_t\}$, together with an ergodicity assumption.

The interiority Assumption (4) may be verified by letting $\theta = (\beta' , \sigma_\varepsilon^2)'$ and $\theta_o = (\beta_o , \sigma_{\varepsilon o}^2)'$. Since the parameter space Θ is assumed not to contain the region $\sigma_\varepsilon^2 \leq 0$ (for the Probit $\sigma_{\varepsilon o}^2 = 1$), it suffices to assume that Θ contains an open neighbourhood of θ_o.

Assumption (5), which requires twice-continuous differentiability on Θ , follows from the fact that the likelihood contributions for both models are continuous functions of the expressions

$$f = \frac{1}{\sigma_\varepsilon} \phi(\frac{k}{\sigma_\varepsilon}) \text{ and } F = \Phi(\frac{K}{\sigma_\varepsilon}) ,$$

which are twice continuously differentiable in k and σ_ε^2 .

Finally, Assumption (6) requiring the non-singularity of the information matrix $L_{\theta\theta}(\theta_o)$), is checked as follows: For the AR(1) Tobit case, one can apply the proof of the negative-definiteness (and hence non-singularity) of $L_{\theta\theta}(\theta_o)$ given by Amemiya (1973), and for the AR(1) Probit case the concavity proof by Pratt (1981). The only modification needed is that, whereas in the i.i.d. cases the strong convergence of $L^T_{\theta\theta}(\theta) = E\lambda^t_{\theta\theta}(\theta)$ follows from the Kolmogorov SLLN, the serial dependence in our case necessitates the SLLN given by Andrews (1986).

One can similarly show that neglecting an <u>error-components structure</u> in the simple <u>panel</u> binomial probit and Tobit cases <u>with no lagged dependent variables</u>, still produces consistent estimates. Consider the stochastic process with the one-factor error-components structure

$$\varepsilon_{it} = \eta_i + \upsilon_{it} \sim N(0, V_\varepsilon) \ , \ V_\varepsilon = \sigma^2_\upsilon \cdot I_{NT} + \sigma^2_\eta \cdot (I_N \otimes i_T i'_T) \ . \tag{9}$$

See Hausman and Taylor (1981). i_T is a T x 1 vector of 1's.
In the terminology introduced above, this process is τ-dependent where $\tau = \max_i \{T_i\}$. This is because all correlations vanish for any time period separation greater than $T_i < \infty$, the number of observations for individual i, and is therefore mixing. The asymptotic results follow when the total number of individuals $N \longrightarrow \infty$. Such an error process will be homoskedastic with variance $(\sigma^2_\eta + \sigma^2_\upsilon)$. Of course as in all examples given here the estimated covariance matrices will be wrong in view of the neglected serial dependence.[3]
Hajivassiliou (1985b) presents an analysis of panel-data models with limited dependent variables, which incorporate state dependence and one-factor unobservable heterogeneity.

<u>Section II</u>

<u>Serial Correlation and Logistic Distributional Assumptions</u>

The consistency proof given in Section I readily suggests difficulties under logistic distributional assumptions on the error ε_t. In exact analogy to (4) and (4'), when ε_t is i.i.d. logistic, we have:

(10) $$f_t = L_t^{y_t} \cdot (1-L_t)^{1-y_t} \quad \text{for Logit ,}$$

and

(10') $$f_t = (1-L_t)^{1-d_t} \cdot \ell_t^{d_t} \quad \text{for L-tobit ,}$$

where $L_t = 1/(1 + \exp(-(y_t - x_t'\beta)/\xi_\varepsilon)$ and $\ell_t = \frac{1}{\xi_\varepsilon} \exp(-(y_t - x_t'\beta))/\xi_\varepsilon)/(1 + \exp(-(y_t - x_t'\beta)/\xi_\varepsilon)^2$. ξ_ε is related to the variance of ε_t by $\sigma_\varepsilon^2 = \xi_\varepsilon^2 \pi^2/3$,

The analogous L-switching model can be obtained by letting the demand and supply shocks be independently and identically distributed extreme-valued with common variance. Then excess demand will be logistically distributed, since the difference of two i.i.d. extreme value r.v.'s will be logistic (see, for example, McFadden (1973)). The Likelihood contribution for this model, which is straightforward to obtain, is given in a previous version of this paper.

We examine the effects of time dependence by considering a stationary AR(1) process for ε_t:

Proposition 1: *Consider the strictly stationary AR(1) process* $\varepsilon_t = \rho\varepsilon_{t-1} + \eta_t$, $|\rho|<1$. *If the innovation η_t is independently identically distributed as logistic,*[4] *the marginals ε_t cannot be logistically distributed. Conversely, for the marginals to be logistically distributed, the innovation cannot follow the logistic law.*

Proof: The characteristic function of a logistically distributed r.v. W with mean 0 and variance = $\xi_W^2\pi^2/3$ can be shown to be:

(11) $$\Psi_W(t) = \pi\, \xi_W\, t \operatorname{cosech}(\pi\, \xi_W\, t),$$

where cosech(u) is the hyperbolic cosecant function $\frac{2}{e^u - e^{-u}}$. Now suppose that ε and ε_{-1} are logistic with mean 0 and scale parameters ξ and ξ_{-1} respectively.

Suppose further that the innovation is also logistic, mean 0 and ξ_η normalized at $1(\Longleftrightarrow \sigma_\eta^2/3)$. By stationarity

$$\xi_\varepsilon = \xi_{\varepsilon_{-1}} \text{ and } \sigma_\varepsilon^2 = \sigma_\eta^2/(1-\rho^2), \text{ so } \xi_\varepsilon^2 = 1/(1-\rho^2).$$

By the independence of η from ε_{-1} that is implied from the assumption of i.i.d. η's, the characteristic functions of the three processes must obey the relation:

$$(12) \qquad \pi\, \xi_\varepsilon t \operatorname{cosech}(\pi\, \xi_\varepsilon t) = \pi\, \xi_\varepsilon \rho\, t \operatorname{cosech}(\pi\, \xi_\varepsilon\, \rho\, t)\, \pi\, \xi_\eta\, t \operatorname{cosech}(\pi\, \xi_\eta\, t)\ .$$

This exploits the result that if X,Y independent with characteristic functions $\Psi_1(t)$ and $\Psi_2(t)$ respectively, then the characteristic function of kX + Y, k=real constant, is equal to $\Psi_1(kt) \cdot \Psi_2(t)$. (See Lukacs (1970)). Using the normalizations and results on the ξ's above, we obtain:

$$(13) \qquad \sinh(\pi\, \rho\, t / \sqrt{1-\rho^2}\,) \cdot \sinh(\pi\, t) = \rho\, \pi\, t \sinh(\pi\, t / \sqrt{1-\rho^2}\,).$$

This relation must hold for all real t and $|\rho|<1$, if ε, ε_{-1} and η are to be logistic with the specified parameters. But this requirement fails to hold in general. For example, when t=1/2, (13) holds trivially for $\rho = 0$ (since sinh(0) = 0). In general, however, (13) is a hyperbolic trigonometric equation that defines a (possibly empty apart from 0) finite set of solutions for ρ, $|\rho|<1$.

<u>Proposition 2</u>: *For ε_t to be distributed as logistic, with $\sigma_\varepsilon^2 = \xi_\varepsilon^2\, \pi^2/3$ and be strictly stationary with i.i.d. innovations, the p.d.f. of the innovation η_t must be characterized by the relation*

$$(14) \qquad \pi\, \xi_\varepsilon\, t \operatorname{cosech}(\pi\, \xi_\varepsilon\, t) = \pi\, \xi_\varepsilon\, \rho\, t \operatorname{cosech}(\pi\, \xi_\varepsilon\, \rho\, t) \cdot \Psi_\eta(t)$$

and the inversion formula for absolutely integrable characteristic functions:

$$(15) \qquad f_\eta(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-itx)\, \Psi_\eta(t)\, dt\ .$$

Proof: The result follows from (11) and the independence of the innovation η_t from ε_{t-1}, implied by the assumption of i.i.d. innovations.

These two propositions show that logistic distributional assumptions are in general incompatible with ARMA-type serial correlation assumptions, in the following sense: If $\rho = 0$, η_t and $\varepsilon_t = 0 + \eta_t$ can (trivially) both be logistic, while if $\rho \neq 0$ ($|\rho|<1$), then $\eta_t = \varepsilon_t - \rho\, \varepsilon_{t-1}$ and the ε's cannot all be logistic. Moreover for ε_t, ε_{t-1} to be logistic, η_t must have a p.d.f. characterized by equations (14) and (15). Only then will the main consistency requirement (Assumption 2 of Theorem 1), that $f(y^*|x, \theta_o)$ be the logistic, hold. As a result, the misspecified ML estimators for Logit and Ltobit that rely on equations (10) and (10'), cannot be shown to be consistent using Theorems 1 and 2, since they do not use a valid conditional density for $y^*|x$ in case the innovations are logistic.

Moreover, Propositions 1 and 2 point out that OC type of estimators, as those proposed by Avery et al. (1983) for the multiperiod Probit and Tobit models, could also not be shown to be consistent in general using our methods, if their logistic counterparts are attempted, since the regression function $E(y_t^*|x_t,\theta)$ is the logistic only under the specific conditions of equations (14) and (15).[5]

It is interesting to examine whether any distribution exists for the innovation η that would satisfy (14) and (15) and yield logistic ε's. That such a distribution exists for $|\rho| = \frac{1}{2}$ is proved by the following proposition:

Proposition 3:

When $|\rho| = \frac{1}{2}$, *the function*

$$(16) \qquad \Psi_\eta(t) = \frac{\pi\xi_\varepsilon t}{\sinh(\pi\xi_\varepsilon t)} \cdot \frac{\sinh(\pi\xi_\varepsilon \rho t)}{\pi\xi_\varepsilon \rho t}$$

is a valid characteristic function, uniquely characterizing the hyperbolic cosecant distribution (Feller (1972)), with probability density function $f(y) = \frac{1}{\pi}\,\mathrm{cosech}(y)$.

Proof: Trivial algebra shows that

$$\Psi_\eta(t) = \frac{1}{\rho}\,\frac{e^{(\rho-1)x}\;(1-e^{-2\rho x})}{(1+e^{-x})\;(1-e^{-x})}\text{, where } x=\pi\xi_e t. \text{ When } |\rho| = \frac{1}{2}\text{,}$$

$$\Psi_\eta(t) = \frac{2}{e^{x/2}(1+e^{-x})} = \frac{2}{(e^{x/2}+e^{-x/2})} \equiv \mathrm{sech}\left(\frac{x}{2}\right).$$

The p.d.f. is derived in Feller (1972) using inversion formula (15) above.

We therefore see that at least when $\rho = \pm\frac{1}{2}$, there exists a distribution for the i.i.d. innovation η in a stationary AR(1) process, such that the marginals are logistic. It is an open question whether such an existence result generalizes to any value of ρ ($|\rho|<1$). Proposition (3) suggests that the distribution for η, if such an η exists, will depend on the value of ρ in a more fundamental way than just through scale or location.

Section III

The Operational Characteristics of Logit and Probit Estimators Under Serial Correlation

Consider the simple binary discrete-choice model with i.i.d. errors. Since in this case the quasi-loglikelihood contribution is differentiable, the key consistency requirement is equivalent to the gradient of the maximand having a zero expectation under the population distribution, at the true parameter vector. See, for example, Amemiya (1973). For the Probit case, the consistency requirement is that

$$(17) \qquad E \frac{y-\Phi}{\Phi(1-\Phi)} \phi \cdot x^j = 0 \qquad j=1,\ldots,k$$

where $\Phi = \Phi(x_t'\beta)$ and $\phi = \phi(x_t'\beta)$; the expectation is carried over the true distribution of ε, conditional on the X's. Denote the true conditional cumulative density of ε evaluated at $x_t'\beta$ by G; expression (17) then becomes

$$(17') \qquad \frac{G-\Phi}{\Phi\cdot(1-\Phi)} \cdot \phi \cdot x^j = 0.$$

This argument shows clearly why the Logit estimator is so close (up to a different scale standardization) under i.i.d. conditions to the probit one: if the errors are logistic, the LHS of expression (17) becomes approximately zero for certain data-parameters configurations, since the logistic G is a good approximation to the cumulative normal Φ, except at the tails of the distribution.

By analogy, one would expect the closeness of probit and logit estimators to carry over to the serially dependent case, since our method of proof shows that it is the marginal distribution of the ε's that will determine the consistency properties. The degree of temporal dependence will in general determine the <u>speed of convergence</u> to the asymptotic limits. Therefore our theoretical results in Section II do not prejudge in any way the ability, under serial correlation, of the logit estimator to still approximate closely the probit results. Given that the quality of the approximation worsens at the tails, an interesting possibility to examine is that the probability of getting observations at the tails goes up as $|\rho|$ rises towards unity[6].

In order to investigate these issues, we perform a Monte-Carlo analysis of the behaviour of the simple (binary) probit and logit estimators when the errors follow AR(1) stochastic processes with: (a) normally and (b) logistically distributed innovations.[7] The first method of analysis is a variant of the technique used by Hausman and McFadden (1984) to examine the power characteristics of Lagrange Multiplier, Wald, Likelihood Ratio and Hausman tests in the Multinomial Logit and Nested Multinomial Logit models of discrete choice. In these Monte Carlo experiments, we employ the single-regressor model:

$$(18) \qquad y_t^* = x_t\beta + \varepsilon_t \,,\; \varepsilon_t = \rho\, \varepsilon_{t-1} + \eta_t \quad E\eta_t = 0 \,,\; E\eta_t^2 = \sigma_\eta^2$$
$$\text{observed } y_t = 1 \quad \text{if } x_t'\beta + \varepsilon_t > 0 \quad t = 1,\ldots,T$$
$$= -1 \quad \text{otherwise}$$

The number of observations is T. Suppose x_t is constructed to take only two values: 2 or -2. Then the observation vector

$(y_t, x_t) \in A = \{(1,2),(1,-2),(-1,2),(-1,-2)\} = \{ a_1 \,,\, a_2 \,,\, a_3 \,,\, a_4 \}$

Denote by n_i the number of observations that are equal to a_i (i=1,...4). Then the Probit estimator of β is given by

$$(19) \qquad b_P = \underset{\beta}{\operatorname{argmax}} \{(n_1+n_4) \log \Phi(2\beta) + (n_2+n_3) \log \Phi(-2\beta)\} \,.$$

and the Logit estimator by

$$(19') \qquad b_L = \underset{\beta}{\operatorname{argmax}} \{(n_1+n_4) \log \frac{1}{1+e^{-2\beta}} + (n_2+n_3) \log \frac{1}{1+e^{2\beta}}\} \,.$$

Using the first order conditions, we obtain the analytic expression:

$$b_P = \frac{1}{2}\Phi^{-1}(r) \ . \tag{20}$$

This is because β lies in the interior of a compact parameter set, and hence $\Phi(2\beta)$ and $\Phi(-2\beta)$ are bounded away from 0. The expression for the variance V_P can be calculated using the second derivative at the optimum ($r = \Phi(2b_p)$).

Correspondingly for the Logit case we obtain

$$b_L = \frac{1}{2}\log\frac{r}{1-r} \ . \tag{20'}$$

The expression for the asymptotic variance of the logit estimator V_L can similarly be calculated using the 2nd derivatives evaluated at the optimum.

Note that these formulae are the <u>analytic</u> (up to machine accuracy for evaluating log(.), $\phi(.)$ and $\Phi(.)$) expressions for the MLE estimates b_P and b_L and of their variances V_P and V_L. Their random properties are induced by their dependence on the ratio $r(=(n_1+n_4)/T)$. Under i.i.d. conditions, as in Hausman and McFadden (1984), (n_1,n_2,n_3,n_4) follows a quadri-nomial distribution, which enables us to obtain the <u>exact</u> distribution (up to machine accuracy) of r and therefore of b_P, b_L, V_P and v_L.

In our case, however, the fact that ε follows a serially correlated AR(1) process means that the distribution of (n_1,n_2,n_3,n_4) conditional on the data vectors $\{y_t,x_t\}$, does not have a tractable analytic characterization. We therefore modify the Hausman and McFadden (1984) procedure and generate a large number N (=1000) of runs of sequences $\{y_t,x_t\}$ of length T, thus obtaining an empirical distribution for r. The x's are generated as i.i.d. drawings from a population $X = 4.B - 2$, where B is a Bernoulli random variable with parameter 0.5. Once a distribution for r is obtained, we proceed to calculate the induced distributions of b_P and b_L. Finally the calculated variances V_P and V_L are compared to the variances of the empirical distributions of the b's. The design of these experiments is described in Table 1a.

A second set of experiments was performed to examine the behaviour of the estimators with multiple exogenous regressors. The design incorporated the model:

(21) $y_t^* = \alpha + \beta x_{1t} + \gamma x_{2t} + \varepsilon_t$, $\varepsilon_t = \rho \varepsilon_{t-1} + \eta_t$, $E\eta_t = 0$, $E\eta^2 = \sigma_\eta^2$

The regressors were generated as

(22) $X_{1t} = 2 B_t - 1 + Z_t$, $X_{2t} = 2 B_t - 1 - 2 Z_t$.

where B_t was a Bernoulli random variable with parameter 0.4 and Z_t exponentially distributed with mean 1. These choices were made in order to introduce some skewness and non-smoothness in the regressors. A correlation between X_1 and X_2 of -0.33 was thus implied. Table 1b summarizes the design of the multiple regressor experiments. Analytic expressions for the estimators are no longer possible; hence iterative optimization algorithms are now employed. Note that in the second set of experiments the initial conditions ε_o were assumed randomly drawn and stationary. For the normality case, we obtain directly that $\varepsilon_o \sim L(\sigma, \sigma_\varepsilon^2)$. The results of Section II, however, explicitly argue against assuming that $\varepsilon_o \sim L(0, \sigma_\varepsilon^2)$, when the innovations are logistic. In these cases, therefore, since the distribution of ε_o is unknown, we started the process at T = -100, thus obtaining an approximately stationary ε_o .

Results

Tables 2 through 4 present the results of the one regressor Monte Carlo experiments, and Tables 5 through 7 the two-regressor results. In the one-regressor experiments the number of replications was N = 1000 whereas the considerable computational burden with multiple regressors restricted N to 100 replications for these experiments. The tabulated numbers are the averages over each set of replications.

The first row in each table contains the i.i.d. case with $\rho=0$. The variance of the marginal error ε was standardized at 4, and 6 different values for ρ were considered (therefore $\sigma_\eta = 2\sqrt{1-\rho^2}$). For each set-up, the tables give the (identified) values of the true parameters. The Logit point- and variance-estimates were appropriately standardized for the $\pi/\sqrt{3}$ factor (see Amemiya (1981)) in order to make them as closely comparable as possible.

The estimates of the estimable quantities $\alpha/\sigma_\varepsilon$, β/σ_ε, and $\gamma/\sigma_\varepsilon$ are denoted by a_P, b_P, c_p for probit, and a_L, b_L, c_L for logit. The asymptotic variances appear in parentheses, while the calculated variances from the empirical distributions of the parameter estimates are given in square brackets. In view of the high number of replications, the empirical variances should be expected to closely estimate the theoretical variances of the estimators.

In Part A of each table the results with normally distributed $(0,\sigma_\eta^2)$ innovations are presented, while in Part B we give the results of the exactly analogous experiments with a logistic $(0,\sigma_\eta^2)$ innovation η. Three sample sizes were tried in order to investigate the asymptotic behaviour of the estimators: T=50 in Tables 2 and 5, T=100 in 3 and 6, and T=1000 in Tables 4 and 7.

The first set of conclusions that can be drawn is that as ρ rises in absolute value towards 1, the bias rises quickly for both the probit and logit estimators. The theoretical discussion above which suggested that logit should still be a close approximation to the probit estimator when the errors truly follow normal (serially correlated or not) processes is validated - the logit point estimates are within 10-15% of the probit ones irrespective of the value of ρ (generally underestimating the true values). What is very impressive is the magnitude of the bias that both estimators exhibit as $|\rho|$ rises. With $\rho = .9$ we get a 10-15% bias, while for $\rho = -.9$ the bias rises to 40%; for $|\rho| = .99$ the estimates are approximately 2-5 times their theoretical values.

This finding can readily be explained by our method of proving consistency. Recall that a basic requirement (Assumption 1) was that the stochastic processes be "mixing". With $|\rho| \longrightarrow 1$, AR(1) approaches the random walk process which has exploding autocovariances and therefore is not mixing - hence the asymptotic independence requirements become harder to hold as $|\rho| \longrightarrow 1$. This is borne out very well by the behaviour of the bias as T rises. With T=100 in Tables 4 and 7, the biases start falling. Even with $|\rho|=.99$ the estimates are only 1-2 times the size of the true coefficients, while with T=1000 the point estimates become quite satisfactory. With i.i.d. errors, probit is off by less than 1% and logit by 10%. More importantly, with $|\rho|=.9$ the biases are now only 10-20%, compared to several orders of magnitude with smaller T's. We thus observe that the rate of convergence rests crucially on ρ; the closer $|\rho|$ is to 1, the higher the sample sizes needed to achieve acceptably low biases.

It is important to note that the situation is not as satisfactory with respect to the standard errors obtained under the false assumption of i.i.d. errors. The consistency of the variance estimator for Probit and (approximate consistency) of the variance estimator for Logit for $\rho = 0$ is validated by observing that the variance estimates differ from the calculated variances of the respective empirical distributions by less than 10% when T=1000 (and by 20% for T=50). However as explained above, once $\rho \neq 0$ these standard errors are inconsistent, the size of the bias rising fast as the degree of serial correlation ($|\rho|$) rises. Moreover the biases <u>remain</u> <u>unacceptably</u> <u>very</u> <u>high</u> even with T=1000. For example, with T=50 the estimated variance of probit underestimates the estimated theoretical variance by 40% when $\rho = .9$, and by 15(!) times for $\rho = -.9$. Even with T as large as 1000, the 40% bias remains for $\rho = .9$, while in the $\rho = -.9$ case we still observe an astonishingly high bias of 9 times the empirical estimates for the variances.[8]

The main aim of the experiments with AR(1) processes with logistic innovations was to examine the importance of the theoretical result (Proposition 1 in Section II) that when the innovation η is logistic, the marginals ε are neither logistic nor normal. Indeed for particular values of the autocorrelation coefficient ρ, the marginal ε may be distributed more closely to a normal than to a logistic random variable, "close" defined by reference to higher order moments.[9] As in the case with normal innovations, we see that again the biases become very large as $|\rho| \longrightarrow 1$, and improve somewhat with larger T's. However, given that the marginals are neither exactly normally nor logistically distributed, neither estimator is consistent. The calculated biases appear almost uniformly worse case by case for both estimators as compared to the first part of each table. Moreover, we now see that generally the logistic seems to perform better as an approximation to the distribution of ε, but not uniformly so. (The logit is consistent, of course, for $\rho = 0$. See especially Tables 5a and 8a where bias is only 2-3%). For example, with $\rho = .5$ the probit performs approximately as well as the logit.

To summarize: Our first finding is that the degree of serial correlation affects crucially the speed of convergence for the consistency properties of the estimators. The biases remain unacceptably high for moderate sample sizes, when ρ is close to 1 in absolute value. Moreover, the precision of the estimators is crucially dependent on $|\rho|$. Secondly, the estimates for the standard errors that falsely assume i.i.d. errors are very badly inconsistent, even for moderate values of ρ; for reliable inferences, therefore, it becomes imperative that one calculates consistent estimators for the standard errors, if usual probit or logit estimators are to be used in the presence of serial correlation. Finally, if logistic innovations are generating the AR stochastic processes, neither probit nor logit is fully appropriate. In general, the logistic distribution performs as a better approximation than the normal, for such a process.

Section IV

Conclusions

This paper analyzed the problem of temporal dependence in limited dependent variables models. A general method to prove consistency under "mixing" dependence conditions was offered and illustrated for the Probit and Tobit models, the stochastic terms assumed AR(1) processes with normal innovations. Time dependence arising because of an errors-components structure in the unobservables was shown to satisfy the "mixing" conditions required by our Theorems of strong consistency. It was further explained how the Theorems could be applied to establish the consistency of the simple disequilibrium switching regressions estimator under such serial dependence conditions.

We also pointed out that special problems arise in proving consistency and asymptotic normality of limited dependent variables models estimators when the errors do not follow normal serially correlated processes. The case of logistic distributional assumptions was used to illustrate the nature of these problems. The point was made that the additivity of ARMA processes does not blend well with logistic distributional assumptions.

A set of Monte Carlo experiments confirmed the consistency results under normal AR(1) conditions, while pointing out that for high levels of serial correlation larger sample sizes are needed ceteris paribus if satisfactorily low levels of bias are to be achieved for the point estimates. Moreover, the logit estimator was found to be a close approximation to the probit even with AR(1) errors. The standard errors, however, of both probit and logit estimators that neglect the serial dependence in the errors were found badly inconsistent, thus arguing that it is imperative to use autocorrelation-consistent standard errors in these cases.

An open problem is the search for estimators that would offer a satisfactory compromise between the tractability of (consistent) inefficient estimators that neglect serial dependence on one hand, and the intractability of the fully efficient Maximum Likelihood estimators that incorporate the time dependence. It seems potentially fruitful to approach the problem by an analysis of the properties of conditional quasi maximum likelihood estimators using the theoretical results presented and referenced to in this paper. We leave this to future work.

Footnotes

[1]Methods of estimating consistently the standard errors for the mis-specified MLE under time-dependence conditions are given in Robinson (1982). The general problem of consistent estimation of covariance matrices under serial correlation is addressed in Newey and West (1986).

[2]Robinson (1982) and Lee (1984a) propose the use of appropriately defined residuals for estimation of and inference on ρ.

[3]Recall the comments on Theorem 1 which pointed out that Assumption (2) required only that we work with the correct conditional density, NOT that we use the right likelihood for the sample $(Z_1,\ldots,Z_T)$. Hence the results in White, etc., that yield Theorem 1 and allowed us in this Section to study the behaviour of misspecified MLE, also suggest here a feasible alternative to full MLE. Specifically, one could consider estimators which evaluate the correct expressions for the densities of a characteristic (under stationarity) period, conditioned on chosen subset of the available information; the resulting <u>Conditional Quasi Maximum Likelihood</u> (CQMLE) estimators can then be shown to be consistent and asymptotically normally distributed. The properties of such estimators are examined in Hajivassiliou (1985a). The analogies to Ruud (1981) and Avery et al. (1983) OC-type of estimators should be clear. The key difference is that CQMLE estimators would work with the correct density conditioned on a chosen subset of the past information, while the OC-type estimators employ the first and second moments of the density, conditional on the chosen information subset.

[4]Strictly stationary processes are characterized by invariant with t joint distributions, while in the case of weak stationarity only the autocovariance functions are invariant. The two coincide for the case of Gaussianity, since then independence implies and is implied by uncorrelatedness. In the absence of Gaussianity we impose a stricter requirement on η_t, namely being i.i.d. rather than just uncorrelated i.d. See Granger and Newbold (1977).

[5]Note that our results have only shown that Theorems 1 and 2 of Section I do not hold for additive, finite ARMA-type processes with logistic innovations, because the marginal errors are not logistically distributed. It should be pointed out that there exist time-dependent processes that have logistic marginals, along the lines of Johnson and Kotz (1972). Such processes, however, are not analogous to ARMA processes, which are widely used in time-series econometric applications.

[6]A second issue, that we do not address here, is the efficiency characteristics of such estimators that neglect the temporal dependence of the errors. From the linear model we know that the inefficiency resulting from falsely assuming i.i.d. errors under serial correlation, rises quickly with the degree of the serial dependence, possibly reaching very high levels (see Malinvaud (1980)).

[7]Note that the problems of ML estimation in limited dependent models with serial dependence, are not specific to AR(•) processes, but arise in the same

way in MA(·) specifications. For example, even assuming the MA(1) process $\epsilon_t = \rho\, \epsilon_{t-1} + \eta_t$, the non-linearities of limited dependent variables models still introduce an uncut integration chain into the likelihood.

[8]An interesting possibility would be to investigate the performance of the Newey and West (1985) estimator for the standard errors in temporal dependence case.

[9]To investigate this possibility empirically, we generated series of 1000 realizations of AR(1) process with (a) normal and (b) logistic innovations. Note that a normal process has a coefficient of kurtosis c_4=0. We calculated the following c_4's, averaged over 1000 replications:

	Normal		Logistic	
	$\tilde{c}_4$ for ϵ_t	$\tilde{c}_4$ for η_t	$\tilde{c}_4$ for ϵ_t	$\tilde{c}_4$ for η_t
ρ = 0	0.0134	0.0134	1.1605	1.1605
ρ = .5	0.0176	0.0134	0.6958	1.1605
ρ = .9	0.0514	0.0134	0.0701	1.1605
ρ = -.5	0.0048	0.0134	0.6802	1.1605
ρ = -.9	0.0543	0.0134	0.0469	1.1605

Note the distinctly closer to 0 values of c_4 in the logistic case when $|\rho|$ is in the 0.5 to 0.9 range.

REFERENCES

Amemiya, T. (1973): "Regression Analysis when the Dependent Variable is Truncated Normal," Econometrica (November).

Amemiya, T. (1981): "Qualitative Response Models: A Survey," Journal of Economic Literature, 19.

Andersen, E.B. (1970): "Asymptotic Properties of Conditional Maximum Likelihood Estimators," Journal of the Royal Statistical Society, B 32.

Andrews, D.W.K. (1986): "Consistency in Nonlinear Econometric Models: A Generic Uniform Law of Large Numbers," Cowles Foundation Discussion Paper, Yale University.

Avery, R.B., Hansen, L.P. and V.J. Hotz (1983): "Multiperiod Probit Models and Orthogonality Condition Estimation," International Economic Review, V. 24, No. 1 (February).

Billingsley, P (1979): Probability and Measure. New York: John Wiley and Sons.

Butler, J.S. and R. Moffit (1982): "A Computationally Efficient Quadrature Procedure for the One-Factor Multinomial Probit Model," Econometrica 50.

Chamberlain, G. (1980): "Analysis of Covariance with Qualitative Data," Review of Economic Studies XLVII.

Fair, R.C. and D.M. Jaffee (1972): "Methods of Estimation for Markets in Disequilibrium," Econometrica, 40 (May).

Feller, W. (1972): An Introduction to Probability Theory, Vol. 2, Second Edition. New York: Wiley.

Finney, D. (1964): Statistical Methods in Bioassay. London: Griffin.

Granger, C.W.J. and P. Newbold (1977): Forecasting Economic Time Series, New York, Academic Press.

Hajivassiliou, V.A. (1984): "A Computationally Efficient Quadrature Procedure for the One Factor Multinomial Probit Model: A Comment," MIT mimeo.

Hajivassiliou, V.A. (1985a): "The Estimation of Certain Non-Linear Models with Autocorrelated Errors by Conditional Quasi Maximum Likelihood Methods," MIT mimeo.

Hajivassiliou, V.A. (1985b): "A Theoretical and Monte-Carlo Comparison of Alternative Estimators of Panel-Data Non-Linear Models with Unobservable Heterogeneity," Yale University mimeo.

Hausman, J.A. (1978): "Specification Tests in Econometrics," Econometrica, 46 (November)

Hausman, J.A. and D.L. McFadden (1984): "Specification Tests for the Multinomial Logit Model," Econometrica, 52 (September).

Hausman, J.A. and W. Taylor (1981): "Panel Data and Unobservable Individual Effects," Econometrica, 49.

Heckman, J.J. (1981a): "Dynamic Discrete Models," in Structural Analysis of Discrete Data with Econometric Applications by Manski, C. and D. McFadden (eds.). Cambridge: MIT Press.

Heckman, J.J. (1981b): "The incidental parameters problem and the problem of initial conditions in estimating a discrete-time, discrete-data stochastic process," in Structural Analysis of Discrete Data with Econometric Applications by Manski, C. and D. McFadden (eds.). Cambridge: MIT Press.

Hoadly, B. (1971): "Asymptotic Properties of Maximum Likelihood Estimators for the Independent Non-identically Distributed Case," Annals of Mathematical Statistics, 42.

Ibragimov, I.A. and Y.V. Linnik (1971): Independent and Stationary Sequences of Random Variables. The Netherlands: Wolters-Noordhoff.

Jennrich, R.I. (1969): "Asymptotic Properties of Non-Linear Least-Squares Estimators," Annals of Mathematical Statistics, 40.

Johnson, N. and S. Kotz (1972): Distributions in Statistics: Continuous Multivariate Distributions. New York: John Wiley and Sons.

Lee, L.-F. (1984): "The Likelihood Function and a Test for Serial Correlations in a Disequilibrium Market Model," Economics Letters, V.14, Nos. 2-3.

Levine, D. (1983): "A Remark on Serial Correlation in Maximum Likelihood," Journal of Econometrics, 23.

Lukacs, E. (1970): Characteristic Functions. New York: Hafner.

Malinvaud, E. (1980): Statistical Methods in Econometrics. Amsterdam: North-Holland.

McFadden, D. (1973): "Conditional Logit Analysis of Qualitative Choice Behaviour," in Frontiers in Econometrics, P. Zarembka (ed.), New York, Academic Press.

Newey, W.K. and J.L. Powell (1985): "Asymmetric Least Squares Estimation and Testing," Princeton University mimeo.

Newey, W.K. and K. West (1986): "A Simple, Positive Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," Discussion Paper #92, Woodrow Wilson School, Princeton University.

Neyman, J. and E.L. Scott (1948): "Consistent Estimates Using Partially Consistent Observations," Econometrica, 16.

Olsen, R.J. (1978): "Note on the Uniqueness of the Maximum Likelihood Estimator for the Tobit Model," Econometrica, 46 (September).

Pratt, J.W. (1981): "Concavity of the Log Likelihood," Journal of the American Statistical Association, 76.

Quandt, R.E. (1981): "Autocorrelated Errors in Simple Disequilibrium Models," Economics Letters, 7.

Robinson, P (1982): "On the Asymptotic Properties of Models Containing Limited Dependent Variables," Econometrica, (January).

Ruud, P. (1981): "Misspecification in Limited Dependent Variable Models," unpublished Ph.D. Dissertation, M.I.T.

White, H. (1980): "A Heteroskedasticity-Consistent Covariance Matrix and a Direct Test for Heteroskedasticity," Econometrica, 48.

White, H. (1984): Asymptotic Theory for Econometricians. New York: Academic Press.

White, H. (1985): "Estimation, Inference, and Specification Analysis." unpublished monograph, University of California at San Diego.

White, H. and I. Domowitz (1984): "Nonlinear Regression with Dependent Observations," Econometrica, Vol. 52, No. 1 (January).

Table 1a
Summary of the Design of the Single-Regressor AR(1) Experiments

Part A: Normal Innovations

$y^*_t = x_t\beta + \varepsilon_t$, $y_t = \text{sign}(y^*_t)$, $X_t = 4\ B_t - 2$, $B_t \sim$ Bernoulli(0.5)

$\varepsilon_t = \rho\ \varepsilon_{t-1} + \eta_t$, $\varepsilon_o = 0$, $\eta_t \sim N(0,\sigma^2_\eta)$

ρ = 0.	T = 1000	N = 1000
= 0.5	= 100	
= 0.9	= 50	
= 0.99		
=−0.5		
=−0.9		
=−0.99		

$\sigma^2_\varepsilon = 4$ * β=0.5

Part B: Logistic Innovations

Difference from Part A: $\eta_t \sim L(0,\sigma^2_\eta)$

*Chosen to match the variance of X.

Table 1b
Summary of the Design of the Multiple-Regressor AR(1) Experiments

Part A: Normal Innovations

$y^*_t = \alpha + x_{1t}\beta + x_{2t}\gamma + \varepsilon_t$, $y_t = \text{sign}(y^*_t)$

$X_{1t} = 2\ B_t - 1 + Z_t$, $X_{2t} = 2\ B_t - 1 - 2\ Z_t$,

$B_t \sim$ Bernoulli(0.4) , $Z_t \sim$ Exponential(1.0)

$\varepsilon_t = \rho\ \varepsilon_{t-1} + \eta_t$, $\varepsilon_o \sim N(0,\sigma^2_\varepsilon)$, $\eta_t \sim N(0,\sigma^2_\eta)$

ρ = 0.	T = 1000	N = 1000
= 0.5	= 100	
= 0.9	= 50	
=−0.5		
=−0.9		

$\sigma^2_\varepsilon = 4$ $\alpha = 0.5$, $\beta = 0.6$, $\gamma = -0.7$

Part B: Logistic Innovations

Differences from Part A: 1. ε_o - see text

2. $\eta_t \sim L(0,\sigma^2_\eta)$

Table 2

Single Regressor Model
1000 Replications
Sample Size = 50, $\sigma_\varepsilon = 2$
Estimand: $\beta/\sigma_\varepsilon = 0.25$

Asymptotic Variances appear in parentheses, Empirical Variances in square brackets

ρ	σ_η	Probit Estimators b_P	Logit Estimators b_L
Part A': Normal Innovations			
0.	2.	0.2546809	0.2279153
		(0.0078335) [0.0095871]	(0.0074114) [0.0080935]
0.5	1.732	0.2586921	0.2316315
		(0.0073960) [0.0098133]	(0.0074647) [0.0083423]
0.9	0.872	0.2801066	0.2523600
		(0.0067280) [0.0175155]	(0.0079193) [0.0151086]
0.99	0.282	0.5222981	0.5049625
		(0.0042975) [0.0888461]	(0.0222583) [0.0994469]
-0.5	1.732	0.2546230	0.2279220
		(0.0074278) [0.0101054]	(0.0074248) [0.0085352]
-0.9	0.872	0.2762809	0.2490192
		(0.0065166) [0.0184784]	(0.0079391) [0.0161192]
-0.99	0.282	0.5188756	0.5015879
		(0.0044086) [0.0895826]	(0.0221658) [0.1001281]
Part B': Logistic Innovations			
0.	2.	0.2867327	0.2573598
		(0.0069834) [0.0093343]	(0.0077530) [0.0080304]
0.5	1.732	0.2758900	0.2474689
		(0.0069461) [0.0102229]	(0.0076515) [0.0087240]
0.9	0.872	0.2843669	0.2563537
		(0.0065253) [0.0174256]	(0.0079880) [0.0151910]
0.99	0.282	0.5135963	0.4961612
		(0.0060355) [0.0895202]	(0.0218496) [0.0998926]
-0.5	1.732	0.2750415	0.2467136
		(0.0074178) [0.0102365]	(0.0076481) [0.0087753]
-0.9	0.872	0.2774820	0.2502179
		(0.0062990) [0.0190782]	(0.0079829) [0.0166531]
-0.99	0.282	0.5013302	0.4827290
		(0.0043638) [0.0865095]	(0.0207470) [0.0957643]

Table 3

Single Regressor Model
1000 Replications
Sample Size = 100, $\sigma_\varepsilon = 2$
Estimand: $\beta/\sigma_\varepsilon = .25$

Asymptotic Variances appear in parentheses, Empirical Variances in square brackets

ρ	σ_η	Probit Estimators b_P	Logit Estimators b_L
Part A': Normal Innovations			
0.	2.	0.2505130 (0.0036511) [0.0045219]	0.2234666 (0.0036188) [0.0037698]
0.5	1.732	0.2526407 (0.0036080) [0.0046652]	0.2254230 (0.0036304) [0.0038902]
0.9	0.872	0.2625294 (0.0034141) [0.0088231]	0.2350084 (0.0037346) [0.0074651]
0.99	0.282	0.3968001 (0.0028994) [0.0591346]	0.3718513 (0.0069293) [0.0618312]
−0.5	1.732	0.2521087 (0.0035017) [0.0048481]	0.2249579 (0.0036300) [0.0040402]
−0.9	0.872	0.2620948 (0.0033492) [0.0079652]	0.2344880 (0.0037199) [0.0067153]
−0.99	0.282	0.4039304 (0.0026555) [0.0576181]	0.3781811 (0.0068326) [0.0594861]
Part B': Logistic Innovations			
0.	2.	0.2798108 (0.0036670) [0.0044693]	0.2502886 (0.0037661) [0.0037874]
0.5	1.732	0.2698661 (0.0037139) [0.0050681]	0.2412417 (0.0037219) [0.0042843]
0.9	0.872	0.2676080 (0.0034606) [0.0092549]	0.2397404 (0.0037697) [0.0078940]
0.99	0.282	0.3934667 (0.0031339) [0.0524056]	0.3669687 (0.0064462) [0.0538629]
−0.5	1.732	0.2681992 (0.0035681) [0.0048206]	0.2396793 (0.0037099) [0.0040667]
−0.9	0.872	0.2618553 (0.0032640) [0.0087898]	0.2343873 (0.0037305) [0.0074295]
−0.99	0.282	0.3757481 (0.0027114) [0.0518419]	0.3495618 (0.0061081) [0.0523566]

Table 4

Single Regressor Model

1000 Replications

Sample Size = 1000, $\sigma_\varepsilon = 2$

Estimand: $\beta/\sigma_\varepsilon = .25$

Asymptotic Variances appear in parentheses, Empirical Variances in square brackets

ρ	σ_η	Probit Estimators b_P	Logit Estimators b_L
Part A': Normal Innovations			
0.	2.	0.2494962 (0.0003573) [0.0004429]	0.2220464 (0.0003565) [0.0003668]
0.5	1.732	0.2495522 (0.0003600) [0.0004796]	0.2221015 (0.0003566) [0.0003969]
0.9	0.872	0.2512702 (0.0003560) [0.0008754]	0.2237135 (0.0003578) [0.0007264]
0.99	0.282	0.2607915 (0.0003460) [0.0055545]	0.2329899 (0.0003682) [0.0046703]
-0.5	1.732	0.2491628 (0.0003594) [0.0004242]	0.2217407 (0.0003563) [0.0003511]
-0.9	0.872	0.2498704 (0.0003596) [0.0008232]	0.2224321 (0.0003571) [0.0006815]
-0.99	0.282	0.2639751 (0.0003357) [0.0052346]	0.2358644 (0.0003694) [0.0044048]
Part B': Logistic Innovations			
0.	2.	0.2807466 (0.0003527) [0.0004433]	0.2506034 (0.0003718) [0.0003734]
0.5	1.732	0.2681206 (0.0003548) [0.0004453]	0.2390359 (0.0003653) [0.0003726]
0.9	0.872	0.2556944 (0.0003465) [0.0008409]	0.2277398 (0.0003598) [0.0006990]
0.99	0.282	0.2634207 (0.0003565) [0.0053963]	0.2353769 (0.0003693) [0.0045436]
-0.5	1.732	0.2681923 (0.0003557) [0.0004650]	0.2391042 (0.0003654) [0.0003893]
-0.9	0.872	0.2550962 (0.0003589) [0.0008320]	0.2271937 (0.0003595) [0.0006918]
-0.99	0.282	0.2671260 (0.0003404) [0.0055972]	0.2387994 (0.0003715) [0.0047244]

Table 5

Multiple Regressor Model
100 Replications
Sample Size = 50, $\sigma_\varepsilon = 2$.

Estimands: $\alpha/\sigma_\varepsilon = 0.250$ $\beta/\sigma_\varepsilon = 0.350$ $\gamma/\sigma_\varepsilon = -0.300$

Asymptotic Variances appear in parentheses, Empirical Variances in square brackets

		Probit Estimators			Logit Estimators		
ρ	σ_η	a_P	b_P	c_P	a_L	b_L	c_L
Part A': Normal Innovations							
0.0	2.000	0.2979541	0.4787046	−0.3607412	0.3171698	0.5133784	−0.3370047
		(0.3752437)	(0.1636032)	(0.0751309)	(0.7249450)	(0.3089917)	(0.1181540)
		[0.3536485]	[0.2328129]	[0.0880862]	[0.4550607]	[0.3094589]	[0.1314675]
0.5	1.732	0.4027455	0.5874178	−0.3528150	0.5142062	0.6469927	−0.3070444
		(0.8110830)	(0.3114668)	(0.1260590)	(1.6281784)	(0.6616594)	(0.2089812)
		[0.6979101]	[0.4599179]	[0.1518096]	[0.9575587]	[0.6467614]	[0.2428799]
0.9	0.872	0.1882999	0.3660746	−0.2180513	0.4009333	0.3979176	−0.1934803
		(0.8037646)	(0.3266774)	(0.1265089)	(2.6834013)	(0.5175002)	(0.1397020)
		[0.4460172]	[0.2754186]	[0.0731708]	[0.8885186]	[0.4044783]	[0.0994552]
−0.5	1.732	0.1989358	0.4707831	−0.4008661	0.1735404	0.4829105	−0.3956323
		(0.1550297)	(0.1516370)	(0.0691430)	(0.2181318)	(0.1890108)	(0.0699833)
		[0.1920410]	[0.1376996]	[0.0604673]	[0.2056911]	[0.1569938]	[0.0709227]
−0.9	0.872	0.2752761	0.6199964	−0.4845482	0.3426579	0.6495772	−0.4441976
		(0.4992085)	(0.2671181)	(0.3552663)	(1.0260156)	(0.4477548)	(0.4058198)
		[0.5020042]	[0.4022010]	[0.2390729]	[0.6720861]	[0.5033030]	[0.3128376]
Part B': Logistic Innovations							
0.0	2.000	0.2361215	0.5127859	−0.3915029	0.2155671	0.5071024	−0.3749333
		(0.3044447)	(0.2122491)	(0.1367528)	(0.3680969)	(0.2269430)	(0.1365875)
		[0.2535107]	[0.1889330]	[0.1250591]	[0.2667837]	[0.2032874]	[0.1332895]
0.5	1.732	0.2987701	0.4751115	−0.4190854	0.3187744	0.4875599	−0.4034841
		(0.4632063)	(0.1681924)	(0.1460335)	(0.8363758)	(0.3002539)	(0.1841566)
		[0.3963428]	[0.2149022]	[0.2001480]	[0.4961355]	[0.2957126]	[0.1851565]
0.9	0.872	0.2193577	0.3926271	−0.3080081	0.3617789	0.4337192	−0.2901406
		(1.3177179)	(0.5272871)	(0.3585558)	(2.8338627)	(0.6523845)	(0.3830668)
		[0.6521304]	[0.4437620]	[0.2819469]	[0.9357767]	[0.5400064]	[0.3337780]
−0.5	1.732	0.2644876	0.4453503	−0.3350017	0.2927487	0.4855190	−0.3101692
		(0.2157883)	(0.1002935)	(0.0622836)	(0.5225454)	(0.2190051)	(0.0968030)
		[0.2441394]	[0.1419021]	[0.0679874]	[0.3667000]	[0.2365054]	[0.1090284]
−0.9	0.872	0.2101201	0.7370919	−0.5194598	0.2840386	0.7837387	−0.4702389
		(0.4318846)	(0.5401904)	(0.2203287)	(0.9803984)	(0.8000970)	(0.2593458)
		[0.4756398]	[0.5903681]	[0.2620252]	[0.6551404]	[0.7057154]	[0.2885114]

Table 6

Multiple Regressor Model
100 Replications
Sample Size = 100, $\sigma_\varepsilon = 2$.

Estimands: $\alpha/\sigma_\varepsilon = 0.250$ $\beta/\sigma_\varepsilon = 0.350$ $\gamma/\sigma_\varepsilon = -0.300$

Asymptotic Variances appear in parentheses, Empirical Variances in square brackets

ρ	σ_η	Probit Estimators a_P	b_P	c_P	Logit Estimators a_L	b_L	c_L
Part A': Normal Innovations							
0.0	2.000	0.2804216	0.4482492	-0.3249289	0.2697961	0.4703084	-0.3115294
		(0.2094765)	(0.0665884)	(0.0448096)	(0.4578514)	(0.1569111)	(0.0753159)
		[0.1835329]	[0,1016263]	[0.0479419]	[0.2470108]	[0.1502758]	[0.0734691]
0.5	1.732	0.2067355	0.3865725	-0.3773745	0.1596524	0.3859812	-0.3766667
		(0.0732832)	(0.0337953)	(0.0299130)	(0.0617103)	(0.0324795)	(0.0286862)
		[0.0680473]	[0.0342120]	[0.0243217]	[0.0597052]	[0.0346128]	[0.0253526]
0.9	0.872	0.2940361	0.5584539	-0.4561726	0.3567047	0.6035393	-0.4446109
		(0.8233115)	(0.1929045)	(0.1313644)	(1.6902391)	(0.3740768)	(0.1833284)
		[0.4400239]	[0.2802235]	[0.1747494]	[0.6185963]	[0.3780730]	[0.2146763]
-0.5	1.732	0.2144688	0.3840804	-0.3436709	0.1662147	0.3807553	-0.3430703
		(0.0504677)	(0.0271202)	(0.0244748)	(0.0442207)	(0.0254154)	(0.0239211)
		[0.0661666]	[0.0306440]	[0.0207843]	[0.0577882]	[0.0304054]	[0.0215513]
-0.9	0.872	0.2233777	0.4241239	-0.3793060	0.1759230	0.4250228	-0.3785834
		(0.0685660)	(0.0462231)	(0.0331292)	(0.0590571)	(0.0460585)	(0.0322781)
		[0.0697586]	[0.0365279]	[0.0261006]	[0.0616444]	[0.0373476]	[0.0272857]
Part B': Logistic Innovations							
0.0	2.000	0.3079636	0.3899298	-0.3014089	0.2529094	0.3926916	-0.3045609
		(0.0769108)	(0.0380799)	(0.0252410)	(0.0682091)	(0.0381073)	(0.0249241)
		[0.0661703]	[0.0301895]	[0.0194712]	[0.0595773]	[0.0313384]	[0.0207946]
0.5	1.732	0.2540955	0.4011839	-0.3490021	0.2207040	0.4178859	-0.3487705
		(0.1633959)	(0.0500816)	(0.0325518)	(0.2792068)	(0.0940404)	(0.0488341)
		[0.1286223]	[0.0679079]	[0.0368260]	[0.1563444]	[0.0935013]	[0.0509674]
0.9	0.872	0.1918136	0.5503719	-0.4372831	0.1884870	0.5691513	-0.4291703
		(0.3995733)	(0.3583272)	(0.1368928)	(0.5376880)	(0.4979070)	(0.1283383)
		[0.1862000]	[0.2539931]	[0.1089083]	[0.2357645]	[0.2985264]	[0.1137256]
-0.5	1.732	0.2941605	0.3813650	-0.3055924	0.2382415	0.3848495	-0.3087197
		(0.0538923)	(0.0217004)	(0.0163817)	(0.0473927)	(0.0212032)	(0.0158952)
		[0.0660843]	[0.0294341]	[0.0187785]	[0.0593328]	[0.0302967]	[0.0200042]
-0.9	0.872	0.2155897	0.4164325	-0.3530223	0.1670964	0.4142644	-0.3553243
		(0.0667821)	(0.0446791)	(0.0303495)	(0.0572291)	(0.0433958)	(0.0326459)
		[0.0665981]	[0.0342591]	[0.0233538]	[0.0592387]	[0.0352098]	[0.0256369]

Table 7

Multiple Regressor Model
100 Replications
Sample Size = 1000, $\sigma_\varepsilon = 2.$

Estimand: $\alpha/\sigma_\varepsilon = 0.250$ $\beta/\sigma_\varepsilon = 0.350$ $\gamma/\sigma_\varepsilon = -0.300$

Asymptotic Variances appear in parentheses, Empirical Variances in square brackets

		Probit Estimators			Logit Estimators		
ρ	σ_η	a_P	b_P	c_P	a_L	b_L	c_L
Part A': Normal Innovations							
0.0	2.000	0.2548684	0.3536004	-0.2999700	0.1944791	0.3565750	-0.3053793
		(0.0054542)	(0.0025778)	(0.0013508)	(0.0044805)	(0.0024489)	(0.0012597)
		[0.0059955]	[0.0025285]	[0.0015745]	[0.0052474]	[0.0025769]	[0.0016946]
0.5	1.732	0.2468281	0.3515892	-0.3013371	0.1872196	0.3554031	-0.3065957
		(0.0091712)	(0.0031453)	(0.0017664)	(0.0074910)	(0.0030410)	(0.0017242)
		[0.0059467]	[0.0025038]	[0.0015584]	[0.0052031]	[0.0025626]	[0.0016801]
0.9	0.872	0.2457574	0.3657070	-0.3108603	0.1864507	0.3693518	-0.3171827
		(0.0222723)	(0.0037143)	(0.0025115)	(0.0177166)	(0.0037615)	(0.0026953)
		[0.0061135]	[0.0026636]	[0.0016804]	[0.0053754]	[0.0027414]	[0.0018311]
-0.5	1.732	0.2555600	0.3531845	-0.2996364	0.1946374	0.3568010	-0.3057265
		(0.0045864)	(0.0026564)	(0.0012994)	(0.0040186)	(0.0026147)	(0.0013498)
		[0.0059243]	[0.0025107]	[0.0015587]	[0.0051862]	[0.0025668]	[0.0016874]
-0.9	0.872	0.2403451	0.3583814	-0.3038275	0.1813444	0.3617486	-0.3088546
		(0.0059132)	(0.0030751)	(0.0020632)	(0.0049905)	(0.0030648)	(0.0020567)
		[0.0059826]	[0.0025431]	[0.0015914]	[0.0052244]	[0.0025970]	[0.0017087]
Part B': Logistic Innovations							
0.0	2.000	0.3249133	0.3432612	-0.2907540	0.2513331	0.3549162	-0.3033996
		(0.0076666)	(0.0028520)	(0.0021910)	(0.0062320)	(0.0026809)	(0.0021261)
		[0.0060898]	[0.0025691]	[0.0016028]	[0.0054663]	[0.0027246]	[0.0018056]
0.5	1.732	0.2967702	0.3619619	-0.3001862	0.2274242	0.3710658	-0.3113770
		(0.0077158)	(0.0037044)	(0.0019586)	(0.0061474)	(0.0034689)	(0.0019021)
		[0.0061112]	[0.0026506]	[0.0016483]	[0.0054418]	[0.0027808]	[0.0018383]
0.9	0.872	0.2451065	0.3553369	-0.3081274	0.1849309	0.3606974	-0.3150963
		(0.0278968)	(0.0033009)	(0.0028432)	(0.0220298)	(0.0034518)	(0.0029727)
		[0.0060701]	[0.0026086]	[0.0016516]	[0.0053557]	[0.0027080]	[0.0018135]
-0.5	1.732	0.2989201	0.3532618	-0.2916504	0.2300225	0.3616917	-0.3022726
		(0.0052170)	(0.0031829)	(0.0019040)	(0.0046099)	(0.0032749)	(0.0019520)
		[0.0059896]	[0.0025556]	[0.0015777]	[0.0053251]	[0.0026714]	[0.0017550]
-0.9	0.872	0.2472638	0.3537663	-0.3048430	0.1873125	0.3572080	-0.3104648
		(0.0057973)	(0.0022346)	(0.0016332)	(0.0047176)	(0.0022593)	(0.0017616)
		[0.0059989]	[0.0025452]	[0.0015975]	[0.0052445]	[0.0025998]	[0.0017240]