International Finance Discussion Papers

Number 126

October 1978

DISTRIBUTED LAG ORDER DETERMINATION

by

Richard Meese

Distributed Lag Order Determination

by Richard Meese*

Introduction

This paper is organized as follows: parameterization problems endemic to time series models are discussed in section one. New approaches to the parameterization problems are summarized and then applied to the problem of simulataneously estimating the length and the coefficients of a distributed lag regression model. The asymptotic properties of the new estimator of the distributed lag model (DLM) are examined in section two, and the small sample properties of the estimator are examined in section three using Monte Carlo experiments.

1. Time Series Parameterization Problems

Considerable economic analysis is carried out using time series data. The large forecasting models of the U.S. economy that help determine macro-policy are notable examples. Hence it is important that econometricians develop and use estimation procedures that are appropriate for time series. An appropriate regression technique for handling this type of data is generalized least squares (GLS), a procedure that dates back to the work of Aitken (1). Consider the generalized linear regression (GR) model:

$$\underset{(T\mathrm{x}1)}{\underline{y}} = \underset{(T\mathrm{x}K)}{X}\underset{(K\mathrm{x}1)}{\underline{\beta}} + \underset{(T\mathrm{x}1)}{\underline{u}}, \text{ with} \tag{1.1}$$

a) X full column rank,

*Economist International Finance Division. John Geweke and Arthur Goldberger made helpful comments on earlier drafts of the manuscript. Any errors which may remain are my own.

b) $E(\underline{u}|X) = \underline{0}$, and

c) $E(\underline{uu}'|X) = \Sigma$, positive definite.

Since Σ is positive definite, so is Σ^{-1}. Let $G'G = \Sigma^{-1}$. After premultiplication of $\underline{y}$ and X by G, the least squares regression of $G\underline{y}$ on GX is best linear unbiased (BLU). Although theoretically eloquent, this procedure is only a paradigm since the researcher rarely knows the disturbance covariance matrix.

Estimation of $\underline{\beta}$ with an unknown Σ matrix has been one focal point of the time series econometric literature for many years. Since an unrestricted Σ matrix contains $T(T+1)/2$ distinct parameters, some restrictions on the autocovariance function of the disturbance process are necessary. The earliest estimators of $\underline{\beta}$ were derived using the assumption that the disturbance process followed a first order autoregression, AR(1), $u(t) = \rho u(t-1) + \varepsilon(t)$. In what follows $\varepsilon(t)$ shall always denote a process which is independent and identically distributed (iid) with zero mean and variance σ^2. For the AR(1) process Σ contains two unknown parameters, ρ and σ^2, and an asymptotically efficient estimator of $\underline{\beta}$ can be obtained using a variety of procedures.[1] Hannan (7) and Amemiya (3) have worked out estimators of $\underline{\beta}$ with less stringent assumptions on the disturbance process. They assume that the disturbance follows an ARMA (p,q) process.

$$\sum_{j=0}^{p} \alpha_j u(t-j) = \sum_{i=0}^{q} \gamma_i \varepsilon(t-i) \quad \text{where} \tag{1.2}$$

a) $\alpha_0 = 1 = \gamma_0$,

b) p and q unknown non-negative integers,

c) the zeros of $\sum_{j=0}^{p} \alpha_j z^j = 0$ (z complex) and $\sum_{i=0}^{q} \gamma_i z^i = 0$ lie outside the unit circle.

Hannan's estimator of $\underline{\beta}$ is developed in the frequency domain while Amemiya's estimator is formulated in the time domain. Both may be interpreted as multistage GLS procedures which require consistent estimation of the parameters of G or Σ as an intermediate step. Since the autocovariance function of the disturbance process is unknown, the Amemiya and Hannan procedures can only be shown to be asymptotically efficient. To prove asymptotic efficiency (and asymptotic normality) of either coefficient estimator, the estimates of the parameters of G or Σ must "improve" as the sample size increases. This requires that the following conditions be satisfied:

(a) The number of parameters characterizing the disturbance process must be allowed to increase without bound.

(b) The number of observations must increase at a faster rate than the number of parameters so that the ratio of parameters to observations tends to zero as each tends to infinity.

Point (a) ensures that the approximation of the true disturbance process improves as the sample size increases, and point (b) ensures that the estimate of the approximation is consistent.

In this paper we shall be concerned with the estimation of the special case of model 1.1 in which the columns of X are successive lagged values of the same variable. Consider the distributed lag model:

$$y(t) = \sum_{s=0}^{M} \beta(s)x(t-s) + \varepsilon(t) \quad \text{with} \tag{1.3}$$

a) M a fixed <u>unknown</u> non-negative integer,

b) $\sum_{s=0}^{M} \beta(s)^2 < \infty$, $\beta(M) \neq 0$,

c) $(x(t), \varepsilon(t))'$ a zero mean jointly covariance stationary process and

d) $E(\varepsilon(t) | x(t-s)) = 0$ for all t and s.

Model 1.3 has an observation matrix X with unknown column dimension. Although models 1.1 and 1.3 have striking dissimilarities, they have a common parameterization problem. Feasible GLS estimators (Hannan and Amemiya) of model 1.1 with disturbance process 1.2 require close attention to points (a) and (b) above. Since M is unknown in model 1.3, a defensible procedure in this context is to expand the length of the fitted distributed lag indefinitely as sample size increases so that specification error is avoided asymptotically. Again, the number of parameters must be allowed to increase without bound as the sample size tends to infinity, while the ratio of parameters to sample size converges to zero.

In practice, the parameterization problem is solved by increasing the dimension of the parameter space deterministically with sample size T. For example, let m be the maximum length of the distributed lag that is to be fit for a given sample size. If a deterministic rule is followed, we choose m as a function of T, m(T), so that $m \to \infty$ as $T \to \infty$ and $\lim_{T\to\infty} \frac{m(T)}{T} = 0$. This guarantees that for sufficiently large T, $m(T) \geq M$, and under-fitting of model 1.3, i.e., $m < M$, is avoided asymptotically. Also, since $\lim_{T\to\infty} \frac{m(T)}{T} = 0$, the estimator of the coefficients of the distributed lag can be shown to have desirable asymptotic properties.[2] The feasible GLS procedures can be made operational by use of the deterministic rule $m = m(T)$ described above. For Hannan efficient estimation of model 1.1, a consistent estimator of the spectral density of the disturbance process can be obtained by expanding the width of the spectral window as a function of the sample size. Amemiya's estimator of model 1.1 is made operational by expanding the length of the residual autoregression as a function of sample size.[3]

Recently, Akaike (2) and Parzen (13, 14, 15) have suggested a new resolution of the type of parameterization problems discussed above. Both authors have suggested methods for choosing the order (length) of an autoregressive process when the order is unknown. Their procedures are similar to regression strategies since one estimates a set of autoregressive models whose length varies from zero to m, (m chosen as a function of the sample size) choosing the order that is best according to some criterion. Akaike's decision rule is based on the principle of maximum likelihood estimation while Parzen's criterion minimizes the one step ahead mean square prediction error. It has been shown (14, p. 14) that for any autoregressive process and large T, the Parzen criterion selects an order that is bounded above by the order chosen using Akaike's criterion. Although this result does not imply that Akaike's decision rule is less useful, we choose to restrict attention to Parzen's criterion, which is called CAT for "criterion autoregressive transfer function."

Parzen's criterion was derived for use in selecting the order of the estimated autoregressive process, but his decision rule can be applied to the problem of estimating M, the length of the distributed lag in model 1.3. One chooses a lag length m* which gives the minimum value of

$$\text{CAT}(i) = \frac{1}{T}\sum_{j=0}^{i} \hat{\sigma}_j^{-2} - \hat{\sigma}_i^{-2}, \quad i=0,\ 1,\ \ldots,\ m, \qquad (1.4)$$

where $\hat{\sigma}_j^2$ is the residual variance from the regression of y(t) on current and j lagged values of the independent variable x(t). The variable T denotes sample size, and m is chosen as a function of T in a manner described below (section two). A rigorous derivation of criterion 1.4 can be found in Parzen (14, pp. 16-20).

The following is a paraphrase of Parzen's rationale for the use of CAT as a method of order estimation. Let $s(t)$ be a zero mean, normal, covariance stationary process with autocovariance function

$$R_s(v) = E(s(t) \cdot s(t+v)),\ v=0, \pm 1, \ldots \quad . \tag{1.5}$$

We assume that $s(t)$ has autoregressive representation,

$$\sum_{j=0}^{\infty} a_\infty(j)s(t-j) = \varepsilon_\infty(t),\ a_\infty(0) \equiv 1. \tag{1.6}$$

Define the m-memory prediction error as

$$\varepsilon_m(t) = s(t) - E[s(t)|s(t-1),\ s(t-2),\ \ldots,\ s(t-m)]. \tag{1.7}$$

The normality of $s(t)$ implies that the m-memory prediction error is linear in past and present $s(t)$,

$$\varepsilon_m(t) = \sum_{j=0}^{m} a_m(j)s(t-j),\quad a_m(0) \equiv 1. \tag{1.8}$$

Since $\varepsilon_m(t)$ is uncorrelated with past values of $s(t)$,

$$E(\varepsilon_m(t) \cdot s(t-k)) = 0,\ k=1,\ \ldots,\ m, \tag{1.9}$$

the m-memory autoregressive coefficients $a_m(j)$, $j=1, \ldots, m$ can be found by solving a set of m Yule-Walker equations, where $a_m(0)$ is defined to be one:

$$\sum_{j=0}^{m} a_m(j)R_s(j-k) = 0, \; k=1, \ldots, m. \tag{1.10}$$

The m-memory prediction error variance is given by

$$\sigma_m^2 = E(\varepsilon_m(t))^2 = \sum_{j=0}^{m} a_m(j)R_s(j), \quad a_m(0) \equiv 1. \tag{1.11}$$

From 1.6 $\varepsilon_\infty(t)$ is the infinite-memory prediction error obtained from the projection of s(t) on its infinite past,

$$\varepsilon_\infty(t) = s(t) - E[s(t) \mid s(t-1), s(t-2), \ldots]. \tag{1.12}$$

Let σ_∞^2 denote the infinite-memory prediction error variance,

$$\sigma_\infty^2 = E(\varepsilon_\infty(t))^2 = \sum_{j=0}^{\infty} a_\infty(j)R_s(j), \quad a_\infty(0) \equiv 1, \tag{1.13}$$

and define the transfer functions

$$g_m(z) = 1 + \sum_{j=1}^{m} a_m(j)z^j$$

$$g_\infty(z) = 1 + \sum_{j=1}^{\infty} a_\infty(j)z^j$$

for z complex, and let

$$\gamma_\infty(e^{iw}) = \sigma_\infty^{-2} g_\infty(e^{iw}), \tag{1.14}$$

$$\hat{\gamma}_m(e^{iw}) = \hat{\sigma}_m^{-2} \hat{g}_m(e^{iw}), \text{ and}$$

$$\hat{\varepsilon}_m(t) = \sum_{j=0}^{m} \hat{a}_m(j)s(t-j).$$

The $\hat{\sigma}_m^2$, $\hat{a}_m(\cdot)$ and $\hat{g}_m(\cdot)$ are consistent estimates of σ_m^2, $a_m(\cdot)$ and $g_m(\cdot)$ respectively which are obtained by solving the sample Yule-Walker equations, where $\hat{a}_m(0)$ is defined to be one:

$$\sum_{j=0}^{m} \hat{a}_m(j)\hat{R}_s(j-k) = 0,\ k=1,\ \ldots,\ m$$

$$\hat{R}_s(v) = \frac{1}{T}\sum_{j=1}^{T-v} s(t)s(t+v),\ \text{and} \tag{1.15}$$

$$\hat{\sigma}_m^2 = \sum_{j=0}^{m} \hat{a}_m(j)\hat{R}_s(j).$$

The idea is to approximate the autoregressive process 1.6, of unknown but possibly infinite order, by a finite order process so as to minimize the one-step ahead mean square prediction error associated with the approximation of $s(t)$ by an AR(m). As a measure of the one-step ahead mean square prediction error Parzen takes (14, p. 19)

$$\begin{aligned} J_m &\equiv E(\hat{\varepsilon}_m(t) - \varepsilon_\infty(t))^2 \\ &= E\int_{-\Pi}^{\Pi} |\hat{\gamma}_m(e^{iw}) - \gamma_\infty(e^{iw})|^2 f(w)\,dw, \end{aligned} \tag{1.16}$$

where $f(w)$ is a spectral density function on $(-\Pi,\Pi)$ given by

$$f(w) = \frac{1}{2\Pi}\sum_{v=-\infty}^{\infty} e^{-ivw} \cdot R_s(v) = \frac{1}{2\Pi}\left(\frac{\sigma_\infty^2}{|g_\infty(e^{iw})|^2}\right). \tag{1.17}$$

For large T Parzen shows (14, pp. 16-20) that approximately,

$$J_m = \frac{1}{T} \sum_{j=0}^{m} \sigma_j^{-2} + (\sigma_\infty^{-2} - \sigma_m^{-2}). \qquad (1.18)$$

This mean square error expression is the sum of two terms, $(\sigma_\infty^{-2} - \sigma_m^{-2})$ representing bias and $\frac{1}{T} \sum_{j=1}^{m} \sigma_j^{-2}$ representing representing variability of $\hat{\gamma}_m$. Because σ_∞^{-2} is not a function of m, to find m*, the optimal order of the AR process, it is sufficient to find the minimum of

$$\mathrm{CAT}(i) = J_i - \sigma_\infty^{-2} = \frac{1}{T} \sum_{j=1}^{i} \sigma_j^{-2} - \sigma_i^{-2},\ i=1, \ldots, m. \qquad (1.19)$$

In practice, when using the CAT criterion to fit an autoregressive model, it is necessary to replace the σ_j^{-2}, j=1, ..., m in formula 1.19 by their consistent estimates. See Parzen (14, pp. 20-23) for further discussion of this point.[4]

Although CAT's theoretical justification is completely different from that of the residual variance criterion, Theil (18, pp. 543-545), the two methods of determining the order of the distributed lag model 1.3 have similarities. The method of selecting the variables to be included in a regression model by choosing the specification with smallest residual variance can be used since the expected value of the residual variance of the erroneous model minus the expectation of the residual variance of the true model is non-negative. On average, one chooses the correct specification of the regression model, but the residual variance criterion produces estimates of the population variance of y given x that are biased downward, and will not choose the correct specification of the regression model if it is <u>not</u> one of the models that is being considered. For fixed m the expression $\frac{1}{T}\sum_{j=0}^{m}\hat{\sigma}_j^{-2} - \hat{\sigma}_m^{-2}$ converges to $\underset{T\to\infty}{\text{plim}}\,(-\hat{\sigma}_m^{-2})$ because the first term of CAT, $\frac{1}{T}\sum_{j=0}^{m}\hat{\sigma}_j^{-2}$ converges to zero as $T\to\infty$ (see section two below). In large samples, minimizing CAT(i), i=1, ..., m is thus similar to choosing the model specification with smallest residual variance, because minimizing $-\hat{\sigma}_m^{-2}$ is equivalent to choosing the model with smallest $\hat{\sigma}_m^2$. Despite the similarity, we produce asymptotic distribution results in the next section which indicate the superiority of CAT over the residual variance criterion.

2. Properties of an estimator of the distributed lag model 1.3 when CAT is used to determine the order of the lag distribution.

Consider the distributed lag model 1.3 with the additional assumptions

e) $\lim_{s\to\infty} (\mathrm{cov}(x(t), x(t-s)) = 0,$

f) $x(t)$ has finite fourth order moments, and

g) $\varepsilon(t)$ is normally distributed with zero mean and unit variance for all t.

It will be convenient to write model 1.3 in matrix notation,

$$\underset{(T\times 1)}{\underline{y}} = \underset{\substack{T\times(M+1)\\(M+1)\times 1}}{X_M\underline{\beta}_M} + \underset{(T\times 1)}{\underline{\varepsilon}} . \tag{1.20}$$

Let m denote the maximum length of the distributed lag that is to be fit for a given T. For m > M partition $X_m = (X_M, X_{m-M})$ and write 1.20 as

$$\underline{y} = \underset{T\times(M+1),\ T\times(m-M)}{(X_M, X_{m-M})} \cdot \begin{pmatrix} \underline{\beta}_M \\ \underline{\beta}_{m-M} \end{pmatrix} + \underline{\varepsilon}, \text{ where } \underline{\beta}_{m-M} = \underline{0}. \tag{1.21}$$

When m < M partition $X_M = (X_m^*, X_{M-m}^*)$ and write 1.20 as

$$\underline{y} = X_m\underline{\beta}_m^* + \underline{v},\ \underline{v} = \underline{\varepsilon} + X_2^*\underline{\beta}_{M-m}^*,\ \underline{\beta}_M = \begin{pmatrix} \underline{\beta}_m^* \\ \underline{\beta}_{M-m}^* \end{pmatrix} . \tag{1.22}$$

Assumptions (c), (e), and (f) are sufficient to show[5]

$$\underset{\substack{T\to\infty \\ \text{(any fixed M)}}}{\text{plim}} \left(\frac{X_M'X_M}{T}\right) = Q(M), \tag{1.23}$$

where Q(M) is an (M+1)x(M+1) matrix with (i,j)th element equal to cov(x(t-i), x(t-j)). When m > M is fixed, define

$$\underset{\substack{T\to\infty \\ \text{(M,m fixed)}}}{\text{plim}} \left\{\frac{1}{T}\begin{pmatrix} X_M'X_M & X_M'X_{m-M} \\ X_{m-M}'X_m & X_{m-M}'X_{m-M}\end{pmatrix}\right\} = \begin{pmatrix} Q(M) & Q(M,m-M) \\ Q(M,m-M)' & Q(m-M)\end{pmatrix}. \tag{1.24}$$

Given that the limit matrix of 1.24 is nonsingular,

$$\underset{\substack{T\to\infty \\ \text{(M,m fixed)}}}{\text{plim}} \left\{\frac{1}{T}\begin{pmatrix} X_M'X_M & X_M'X_{m-M} \\ X_{m-M}'X_M & X_{m-M}'X_{m-M}\end{pmatrix}\right\}^{-1} \tag{1.25}$$

exists since the elements of the inverse of a matrix are continuous functions of the elements of the matrix itself, (18, p. 363) Similarly, when m < M define

$$\underset{\substack{T\to\infty \\ \text{(m,M fixed)}}}{\text{plim}} \left\{\frac{1}{T}\begin{pmatrix} X_m^{*\prime}X_m^* & X_m^{*\prime}X_{M-m}^* \\ X_{M-m}^{*\prime}X_m^* & X_{M-m}^{*\prime}X_{M-m}^*\end{pmatrix}\right\} = \begin{pmatrix} Q^*(m) & Q^*(m,M-m) \\ Q^*(m,M-m)' & Q^*(M-m)\end{pmatrix}. \tag{1.26}$$

It is also true that limit matrix of 1.26 has an inverse by the same argument given above, (18, p. 363). In what follows we will be interested in expanding the maximum length of the fitted distributed lag indefinitely as $T \to \infty$. Suppose we let $m = m(T)$ so that $m(T) \to \infty$ as $T \to \infty$, and $\lim_{T\to\infty} \frac{m(T)}{T} = 0$. Clearly, for T sufficiently large, $m(T) \geq M$. Now consider the sequence of $(M+1) \times (M+1)$ matrices for any sample size T; $\frac{1}{T}(X_M'X_M) \geq$

$$\frac{1}{T}(X_M'P_1X_M) \geq \frac{1}{T}(X_M'P_2X_M) \geq \cdots \geq \frac{1}{T}(X_M'P_{m(T)-M}X_M') > 0, \tag{1.27}$$

where $\geq$ and $>$ denote matrix ordering and $P_i = (I - X_i(X_i'X_i)^{-1}X_i')$, $i = 1, \ldots, m(T)-M$, where X_i is the $T \times i$ matrix whose j^{th} column, $j = 1, \ldots i$ is composed of $x(t)$ lagged $M + j$ times. The inverses of the matrices in 1.27 also form a monotone sequence,

$$T(X_M'X_M)^{-1} \leq T(X_M'P_1X_M)^{-1} \leq \cdots \leq T(X_M'P_{m(T)-M}X_M)^{-1}. \tag{1.28}$$ [6]

Finally, we define $\hat{\sigma}_m^2$ as the error sum of squares $\underline{\hat{\varepsilon}}_m'\underline{\hat{\varepsilon}}_m$ from a regression of $y(t)$ on $x(t), x(t-1), \ldots, x(t-m)$, $t = 1, \ldots, T$ divided by $T-m-1$.

For any <u>fixed</u> m and iid $\varepsilon(t)$, $\operatorname{plim}_{T\to\infty} (\frac{1}{T} X_m' \underline{\varepsilon}) = \underline{0}$, (4, pp. 23-24).[7] When $m < M$, $\operatorname{plim}_{T\to\infty} (\hat{\sigma}_m^2) =$

$$\lim_{T\to\infty} (T/(T-m-1)) \cdot \operatorname{plim}_{T\to\infty} \frac{1}{T}[\underline{\beta}_M'X_M'P_m^*X_M\underline{\beta}_M + 2\underline{\beta}_M'X_M'P_m^*\underline{\varepsilon} + \underline{\varepsilon}'P_m^*\underline{\varepsilon}], \tag{1.29}$$

where $P_m^* = (I - X_m^*(X_m^{*\prime}X_m^*)^{-1}X_m^{*\prime})$.

$$\operatorname*{plim}_{T\to\infty} (\hat{\sigma}^2_m) = 1 \cdot \left[\underline{\beta}'_M \operatorname*{plim}_{T\to\infty} \left(\frac{X'_M P^*_m X_M}{T}\right)\underline{\beta}_M \right. +$$

$$2\underline{\beta}'_M \operatorname*{plim}_{T\to\infty} \left[\frac{X'_M \varepsilon}{T} - \left(\frac{X'_M X^*_m}{T}\right)\left(\frac{X^{*\prime}_m X^*_m}{T}\right)^{-1} \left(\frac{X^{*\prime}_m \underline{\varepsilon}}{T}\right)\right] + \tag{1.30}$$

$$\operatorname*{plim}_{T\to\infty} \left(\frac{\underline{\varepsilon}'\underline{\varepsilon}}{T}\right) - \operatorname*{plim}_{T\to\infty} \left[\left(\frac{X^{*\prime}_m \underline{\varepsilon}}{T}\right)' \left(\frac{X^{*\prime}_m X^*_m}{T}\right)^{-1} \left(\frac{X^{*\prime}_m \underline{\varepsilon}}{T}\right)\right] =$$

$$\underline{\beta}'_M \operatorname*{plim}_{T\to\infty} \left(\frac{X'_M P^*_m X_M}{T}\right) \underline{\beta}_M + \sigma^2 > \sigma^2.$$ [8]

For $m \geq M$, $\operatorname*{plim}_{T\to\infty} \hat{\sigma}^2_m = \sigma^2$ since $X'_M(I - X_m(X'_m X_m)^{-1} X_m) = 0$ when $m \geq M$. Therefore, $\hat{\sigma}^2_m$ is a consistent estimator of σ^2 when $m \geq M$, m fixed. For $m < M$, $\operatorname*{plim}_{T\to\infty} \hat{\sigma}^2_m > \sigma^2$. Now consider the case $m = m(T)$:

$$\operatorname*{plim}_{T\to\infty} \hat{\sigma}^2_{m(T)} = \operatorname*{plim}_{T\to\infty} \left(\frac{\underline{\varepsilon}'\underline{\varepsilon}}{T-m(T)-1}\right) - \operatorname*{plim}_{T\to\infty} \left(\frac{\underline{\varepsilon}' X_{m(T)} (X'_{m(T)} X_{m(T)})^{-1} X'_{m(T)} \underline{\varepsilon}}{T-m(T)-1}\right). \tag{1.31}$$

Let $N_{m(T)} = X_{m(T)}(X'_{m(T)} X_{m(T)})^{-1} X'_{m(T)}$. Then from 1.31,

$$\operatorname*{plim}_{T\to\infty} \hat{\sigma}^2_{m(T)} = \sigma^2 - \lim_{T\to\infty} \left(\frac{T}{T-m(T)-1}\right) \operatorname*{plim}_{T\to\infty} \left(\frac{\underline{\varepsilon}' N_{m(T)} \underline{\varepsilon}}{T}\right). \tag{1.32}$$

Observe that

$$\left(\frac{\varepsilon' N_{m(T)}\varepsilon}{T}\right) \geq 0 \text{ for all T and m(T), and that}$$

$$E\left(\frac{\varepsilon' N_{M(T)}\varepsilon}{T}\right) = \frac{\sigma^2 m(T)}{T}. \tag{1.33}$$

Clearly, $\lim_{T\to\infty} E\left(\frac{\varepsilon' N_{m(T)}\varepsilon}{T}\right) = \lim_{T\to\infty} \sigma^2\left(\frac{m(T)}{T}\right) = 0$, from which

$\operatorname{plim}_{T\to\infty}\left(\frac{\varepsilon' N_{m(T)}\varepsilon}{T}\right) = 0.$[9] Therefore, $\hat{\sigma}^2_{m(T)}$ is a consistent estimator of σ^2 provided $\lim_{T\to\infty} \frac{m(T)}{T} = 0$.

We are now ready to examine the large sample properties of CAT as a model selection criterion. For sufficiently large T, $m(T) \geq M$ so without loss of generality, we need only examine the case for which $m(T) \geq M$. Consider the probability limit of CAT(m(T)),

$$\operatorname{plim}_{T\to\infty} CAT(m(T)) = \operatorname{plim}_{T\to\infty}\left(\frac{1}{T}\sum_{j=0}^{m(T)} \hat{\sigma}_j^{-2} - \hat{\sigma}_{m(T)}^{-2}\right) =$$

$$\operatorname{plim}_{T\to\infty} \frac{1}{T}\sum_{j=0}^{M} \hat{\sigma}_j^{-2} + \operatorname{plim}_{T\to\infty} \frac{1}{T}\sum_{j=M+1}^{m(T)} \hat{\sigma}_j^{-2} - \operatorname{plim}_{T\to\infty} \hat{\sigma}_{m(T)}^{-2}. \tag{1.34}$$

We shall analyze each part separately:

$$\operatorname{plim}_{T\to\infty} \frac{1}{T}\sum_{j=0}^{M} \hat{\sigma}_j^{-2} = \lim_{T\to\infty}\left(\frac{1}{T}\right)\sum_{j=0}^{M} \operatorname{plim}_{T\to\infty} \hat{\sigma}_j^{-2} \leq \lim\left(\frac{1}{T}\right)(M+1)\sigma^{-2} = 0, \tag{1.35}$$

since for $j < M$, $\underset{T\to\infty}{\text{plim}}\ \hat{\sigma}_j^{-2} < \sigma^{-2}$. Therefore, $\underset{T\to\infty}{\text{plim}}\ \frac{1}{T}\sum_{j=0}^{M}\hat{\sigma}_j^{-2} = 0$. We have already established that $\underset{T\to\infty}{\text{plim}}\ (-\hat{\sigma}^{2}_{m(T)}) = -\sigma^{-2}$, so there is one part of 1.34 left to consider. Now

$$E\left(\frac{1}{T}\sum_{j=M+1}^{m(T)}\hat{\sigma}_j^{-2}\right) = \frac{1}{T}\sum_{j=M+1}^{m(T)} E(\hat{\sigma}_j^{-2}) = \frac{1}{T}\sum_{j=M+1}^{m(T)}\frac{T-j-1}{\sigma^2(T-j-3)}, \tag{1.36}$$

since $\hat{\underline{\varepsilon}}'_j\,\hat{\underline{\varepsilon}}_j/\sigma^2$ is distributed as a $\chi^2(T-j-1)$ for $j=M+1, \ldots, m(T)$, and using the definition of expectation, the reciprocal of a $\chi^2(t-j-1)$ variate can be shown to be $(\frac{1}{T-j-3})$. The last term on the right hand side of 1.36 is bounded above.

$$\frac{1}{T}\sum_{j=M+1}^{m(T)}\frac{T-j-1}{\sigma^2(T-j-3)} \leq \frac{m(T)-M}{T\sigma^2}\left(\frac{T-m(T)-1}{T-m(T)-3}\right). \tag{1.37}$$

The limit as $T\to\infty$ of the right hand side of 1.37 is equal to zero, so $\underset{T\to\infty}{\text{plim}}\ \frac{1}{T}\sum_{j=M+1}^{m(T)}\hat{\sigma}_j^{-2} = 0$ by lemma 1 in footnote 9. Returning to our original problem 1.34, we have

$$\underset{T\to\infty}{\text{plim}}\ \text{CAT}(m(T)) = \underset{T\to\infty}{\text{plim}}\left(\frac{1}{T}\sum_{j=0}^{m(T)}\hat{\sigma}_j^{-2} - \hat{\sigma}^{-2}_{m(T)}\right) = -\sigma^{-2}. \tag{1.38}$$

Provided $m(T)$ increases at a slower rate then T, i.e., $\lim_{T\to\infty}\frac{m(T)}{T} = 0$, CAT(m) is a consistent estimator of $-\hat{\sigma}^{-2}$.[10]

Define m* as the lag length which gives the minimum value of

$$CAT(i) = \frac{1}{T}\sum_{j=0}^{i} \hat{\sigma}_j^{-2} - \hat{\sigma}_i^{-2} \qquad , \ i=0,\ 1,\ \ldots,\ m(T).$$

Proposition 1: The $\lim_{T\to\infty} \text{Prob}(m^* \geq M) = 1$, i.e., the probability that minimizing CAT(i), i=0, 1, ..., m(T) results in specification error, goes to zero as $T\to\infty$.

Proof: Suppose we look at a finite subcollection of the set {i=0, 1, 2, ..., m(T)} that does not contain any integers greater than or equal to M. Let $\tilde{i}$ = {0, 1, 2, ..., M-1}.

Then $\text{plim}_{T\to\infty} [\min_{\tilde{i}} CAT(i)] = \min_{\tilde{i}}(\text{plim}_{T\to\infty} CAT(i))$ since the min function does not depend upon the sample size T, and $\text{plim}_{T\to\infty} CAT(i)$ exists by virtue of 1.30. The $\min_{\tilde{i}} (\text{plim}_{T\to\infty} CAT(i)) > -\sigma^{-2}$ from 1.30.

If we look at a different finite subcollection of the i's that contains at least one integer greater than or equal to M, i* = {non-negative intergers < M, and at least one integer $\geq$ M}, then

$$\text{plim}_{T\to\infty} (\min_{i^*} CAT(i)) = \min_{i^*} (\text{plim}_{T\to\infty} CAT(i)) = -\sigma^{-2}.$$

Since $m(T)\to\infty$ as $T\to\infty$, for sufficiently large T we have $m(T)\geq M$ and there exists a finite subcollection of {i=0, 1, ..., m(T)} that contains at least one integer greater than or equal to M as $T\to\infty$. Therefore, $\lim_{T\to\infty} \text{Prob}(m^* \geq M) = 1$.

Proposition 2: Let k be a fixed positive integer. Then

$T \cdot (CAT(M+k) - CAT(M))$ converges in distribution to
$\frac{1}{\sigma^2}(2k-\chi^2(k))$ as $T\to\infty$.[11]

Proof: $T(CAT(M+k) - CAT(M)) = \sum_{j=M+1}^{M+k} \hat{\sigma}_j^{-2} + T\hat{\sigma}_M^{-2} - T\hat{\sigma}_{M+k}^{-2}$

$$= \sum_{j=M+1}^{M+k} \hat{\sigma}_j^{-2} + \frac{T(T-M-1)}{(\hat{\underline{\varepsilon}}'_M\hat{\underline{\varepsilon}}_M)} - \frac{T(T-M-k-1)}{(\hat{\underline{\varepsilon}}'_{M+k}\hat{\underline{\varepsilon}}_{M+k})} \tag{1.39}$$

$$= \sum_{j=M+1}^{M+k} \hat{\sigma}_j^{-2} + \frac{k\left(\frac{\hat{\underline{\varepsilon}}'_{M+k}\hat{\underline{\varepsilon}}_{M+k}}{T}\right) - \left(\frac{T-M-k-1}{T}\right)\left(\hat{\underline{\varepsilon}}'_M\hat{\underline{\varepsilon}}_M - \hat{\underline{\varepsilon}}'_{M+k}\hat{\underline{\varepsilon}}_{M+k}\right)}{\left(\frac{\hat{\underline{\varepsilon}}'_M\hat{\underline{\varepsilon}}_M}{T}\right)\left(\frac{\hat{\underline{\varepsilon}}'_{M+k}\hat{\underline{\varepsilon}}_{M+k}}{T}\right)}.$$

Because $\left(\frac{\hat{\underline{\varepsilon}}'_{M+i}\hat{\underline{\varepsilon}}_{M+i}}{T}\right)$ is a consistent estimator of σ^2 for $i=1, \ldots,$ $k < \infty$, and $\frac{1}{\sigma^2}(\hat{\underline{\varepsilon}}'_M\hat{\underline{\varepsilon}}_M - \hat{\underline{\varepsilon}}'_{M+k}\hat{\underline{\varepsilon}}_{M+k})$ has a $\chi^2(k)$ distribution independent of T, expression 1.39 converges in distribution to

$$k\sigma^{-2} + \left(\frac{k\sigma^2 - \sigma^2\chi^2(k)}{\sigma^4}\right) = \frac{1}{\sigma^2}(2k - \chi^2(k)).$$

Proposition 3:

$$\text{Cov}(\hat{\underline{\varepsilon}}'_{M+K}\hat{\underline{\varepsilon}}_{M+K} - \hat{\underline{\varepsilon}}'_{M+i}\hat{\underline{\varepsilon}}_{M+i},\quad \hat{\underline{\varepsilon}}'_{M+i}\hat{\underline{\varepsilon}}_{M+i} - \hat{\underline{\varepsilon}}'_{M+j}\hat{\underline{\varepsilon}}_{M+j}) = 0,\ 0 \le j < i \le K.$$

Proof: For $0 \leq j < i \leq K$, it is true that $(\underline{\hat{\varepsilon}}'_{M+j}\,\underline{\hat{\varepsilon}}_{M+j} - \underline{\hat{\varepsilon}}'_{M+i}\underline{\hat{\varepsilon}}_{M+i})$ and $\underline{\hat{\varepsilon}}'_{M+i}\,\underline{\hat{\varepsilon}}_{M+i}$ are independent (18, pp. 84-85, 139). This implies

$$\mathrm{cov}(\underline{\hat{\varepsilon}}'_{M+j}\,\underline{\hat{\varepsilon}}_{M+j} - \underline{\hat{\varepsilon}}'_{M+i}\,\underline{\hat{\varepsilon}}_{M+i},\ \underline{\hat{\varepsilon}}'_{M+i}\underline{\hat{\varepsilon}}_{M+i}) = 0,$$

whence

$$\mathrm{cov}(\underline{\hat{\varepsilon}}'_{M+j}\,\underline{\hat{\varepsilon}}_{M+j},\ \underline{\hat{\varepsilon}}'_{M+i}\underline{\hat{\varepsilon}}_{M+i}) = \mathrm{Var}(\underline{\hat{\varepsilon}}'_{M+i}\underline{\hat{\varepsilon}}_{M+i}) = 2(T-M-i-1)\sigma^4. \tag{1.40}$$

The second line follows from the fact that $\left(\dfrac{\underline{\hat{\varepsilon}}'_{M+i}\underline{\hat{\varepsilon}}_{M+i}}{\sigma^2}\right)$ has a chi-square distribution with (T-M-i-1) degrees of freedom, and the variance of a chi-square variate is equal to twice the number of degrees of freedom. The $\mathrm{cov}(\underline{\hat{\varepsilon}}'_{M+K}\,\underline{\hat{\varepsilon}}_{M+K} - \hat{\varepsilon}'_{M+i}\,\hat{\varepsilon}_{M+i},\ \hat{\varepsilon}'_{M+i}\,\hat{\varepsilon}_{M+i} - \hat{\varepsilon}'_{M+j}\,\hat{\varepsilon}_{M+j}) = 0$ since for $0 \leq j < i \leq K$, and using 1.40 above,

$$\mathrm{cov}(\underline{\hat{\varepsilon}}'_{M+K}\,\underline{\hat{\varepsilon}}_{M+K} - \underline{\hat{\varepsilon}}'_{M+i}\underline{\hat{\varepsilon}}_{M+i},\ \underline{\hat{\varepsilon}}'_{M+i}\,\underline{\hat{\varepsilon}}_{M+i} - \underline{\hat{\varepsilon}}'_{M+j}\,\underline{\hat{\varepsilon}}_{M+j}) = \tag{1.41}$$

$$2(T-M-K-1)\sigma^4 - 2(T-M-K-1)\sigma^4 + 2(T-M-i-1)\sigma^4 - 2(T-M-i-1)\sigma^4 = 0.$$

Using propositions 2 and 3 we can now calculate the $\lim_{T\to\infty} \mathrm{Prob}(m^* = M+j)$, $j=0, 1, \ldots, k$. It has already been established that $\lim_{T\to\infty} \mathrm{Prob}(m^* = M-j) = 0$, $j=1, \ldots, M$. Define the following:

$$S_j \equiv 2\cdot j + (\underline{\hat{\varepsilon}}'_{M+j}\underline{\hat{\varepsilon}}_{M+j} - \underline{\hat{\varepsilon}}'_M\underline{\hat{\varepsilon}}_M)/\sigma^2,\ 0 < j < \infty,$$

$$S_o \equiv 0. \tag{1.42}$$

Observe that $S_j \sim 2\cdot j - \chi^2(j)$ and $S_j = S_{j-1} + 2 - w_j$, where the w_j's are distributed as independent $\chi^2(1)$ variates. Define

$$\rho_j \equiv \mathrm{Prob}(Z_j > 0 \mid Z_{j-1} > 0), \quad j=1,2, \ldots \quad \text{and}$$

$$\rho_i^* \equiv \mathrm{Prob}(-Z_i > 0 \mid -Z_{i-1} > 0), \quad i=1, \ldots, j, \tag{1.43}$$

where $Z_j \sim 2\cdot j - \chi^2(j)$, $Z_o \equiv 0$, $Z_j = Z_{j-1} + y_j$, $y_j = 2 - \chi^2(1)$ and y_j is independent of $Z_{j-1} = \sum_{\ell=1}^{j-1} y_\ell$. Now

$$\lim_{T\to\infty} \mathrm{Prob}(m^* = M+j) = \mathrm{Prob}(-S_j > 0,\ S_1 - S_j > 0, \ldots,$$

$$S_{j-1} - S_j > 0,\ S_{j+1} - S_j > 0,\ S_{j+2} - S_j > 0, \ldots) \tag{1.44}$$

$$= \mathrm{Prob}(-S_j > 0, \ldots, S_{j-1} - S_j > 0) \cdot \mathrm{Prob}(S_{j+1} - S_j > 0, \ldots),$$

since the random variables $(S_i - S_j)$, $0 \le i \le j - 1$ are independent of $(S_n - S_j)$, $n \ge j + 1$ by proposition 3. Note that

$$\mathrm{Prob}(-S_j > 0, \ldots, S_{j-1} - S_j > 0) =$$

$$\mathrm{Prob}(S_{j-1} - S_j > 0) \cdot \mathrm{Prob}(S_{j-2} - S_j > 0 \mid S_{j-1} - S_j > 0) \cdot$$

$$\mathrm{Prob}(S_{j-3} - S_j \mid S_{j-2} - S_j > 0,\ S_{j-1} - S_j > 0) \cdot \ldots \tag{1.45}$$

$$\cdot\ \mathrm{Prob}(S_{j-i} - S_j > 0 \mid S_{j-i+1} - S_j > 0, \ldots, S_{j-1} - S_j > 0).$$[12]

Given $1 \le i \le j$ observe that

$$\text{Prob}(S_{j-i} - S_j > 0 | S_{j-i+1} - S_j > 0, \ldots, S_{j-1} - S_j > 0) =$$

$$\text{Prob}(-Z_i > 0 | -Z_{i-1} > 0, \ldots, -Z_1 > 0) = \tag{1.46}$$

$$\text{Prob}(-Z_{i-1} - y_i > 0 | -Z_{i-1} > 0, \ldots, -Z_1 > 0)$$

$$= \text{Prob}(-Z_{i-1} - y_i > 0 | -Z_{i-1} > 0),$$

because y_i is independent of Z_{i-n}, n=1, ..., i-1. Therefore,

$$\text{Prob}(-S_j > 0, S_1 - S_j > 0, \ldots, S_{j-1} - S_j > 0) =$$

$$\prod_{i=1}^{j} \text{Prob}(S_{j-i} - S_j > 0 | S_{j-i+1} - S_j > 0) = \prod_{i=1}^{j} \rho_i^*. \tag{1.47}$$

Similiarly, one can show

$$\text{Prob}(S_{j+1} - S_j > 0, S_{j+2} - S_j > 0, \ldots) = \prod_{i=1}^{\infty} \rho_i. \tag{1.48}$$

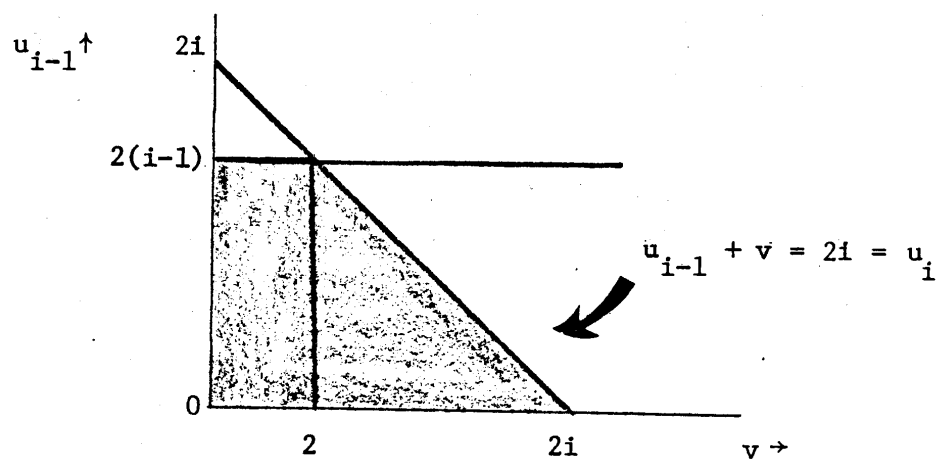
Putting the two results together,

$$\lim_{T\to\infty} \text{Prob}(m^* = M+j) = \prod_{i=1}^{j} \rho_i^* \prod_{i=1}^{\infty} \rho_i. \tag{1.49}$$

Inorder to calculate the limiting probability that m* is equal to M+j, we need to approximate expression 1.49. To accomplish this end let $u_{i-1} \sim \chi^2(i-1)$, $v \sim \chi^2(1)$ independent of u_{i-1}, and for $i \geq 2$ note that

$$\begin{aligned}\rho_i &= \text{Prob}(Z_i > 0 \mid Z_{i-1} > 0) \\ &= \text{Prob}(2i - u_{i-1} - v > 0 \mid 2(i-1) - u_{i-1} > 0) \\ &= \text{Prob}(u_{i-1} + v < 2i \mid u_{i-1} < 2(i-1)) \\ &= \text{Prob}(u_i < 2i \mid u_{i-1} < 2(i-1)).^{13}\end{aligned} \tag{1.50}$$

Figure 1



The shaded area of figure 1 represents the set of u_{i-1} and v such that $u_{i-1} + v < 2i$ and $u_{i-1} < 2(i-1)$. Let $F_{u_i}(\cdot)$, $F_{u_{i-1}}(\cdot)$, and $F_v(\cdot)$ denote the cumulative distribution functions (cdf) of u_i, u_{i-1} and v and let f_{u_i}, $f_{u_{i-1}}$, and f_v denote the corresponding probability density functions (pdf). It is clear from figure 1.1 that there are several representations of ρ_i in terms of these cdf and pdf. Since the ρ_i must be calculated numerically, we have chosen the following expression for ρ_i to minimize computational cost:

$$\rho_i = \frac{F_{u_i}(2i) - \int_0^2 (F_{u_{i-1}}(2i-v) - F_{u_{i-1}}(2(i-1))\, f_v(v)dv}{F_{u_{i-1}}(2(i-1))}$$

$$= F_v(2) + \left(\frac{F_{u_i}(2i) - \int_0^2 F_{u_{i-1}}(2i-v) f_v(v) dv}{F_{u_{i-1}}(2(i-1))} \right). \qquad (1.51)$$

Numerical integration of the second term in the numerator of 1.51 was carried out as follows. The closed interval [0,2] was divided into 20,000 disjoint intervals, and the value of $f_v(v)$, $0 < v < 2$ was calculated using the approximation

$$f_v(v) = \text{Prob}(\chi^2(1) \leq v + .0001) - \text{Prob}(\chi^2(1) \leq v - .0001). \qquad (1.52)$$

A similar procedure is used to find ρ_i^*, $i \leq j$:

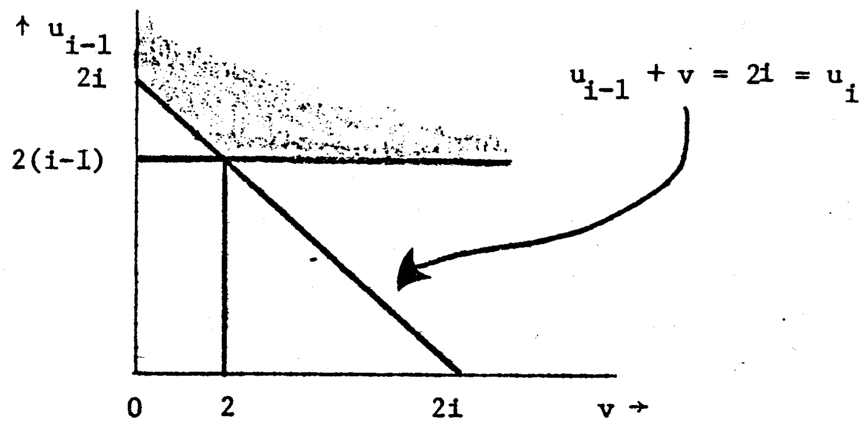
$$\rho_i^* = \text{Prob}(-Z_i > 0 | -Z_{i-1} > 0)$$

$$= \text{Prob}(2i - u_{i-1} - v < 0 | 2(i-1) - u_{i-1} < 0) \qquad (1.53)$$

$$= \text{Prob}(u_{i-1} + v > 2i | u_{i-1} > 2(i-1))$$

$$= \text{Prob}(u_i > 2i | u_{i-1} > 2(i-1)).$$

Figure 2



The shaded area of figure 2 represents the set of u_{i-1} and v such that $u_{i-1} + v > 2i$ and $u_{i-1} > 2(i-1)$.

$$\rho_i^* = \frac{1 - F_{u_{i-1}}(2(i-1)) - \int_0^2 (F_{u_{i-1}}(2i-v) - F_{u_{i-1}}(2(i-1)))f_v(v)dv}{1 - F_{u_{i-1}}(2(i-1))}$$

$$= \frac{1 + F_{u_{i-1}}(2(i-1))F_v(2) - \int_0^2 F_{u_{i-1}}(2i-v)f_v(v)dv}{1 - F_{u_{i-1}}(2(i-1))} \qquad (1.54)$$

Table 2.1 summarizes the results of these calculations.

Table 2.1
Limiting Probabilities that $m^*=M+j$, $j=0, \ldots, 14$.

$M+j$	$\lim_{T\to\infty} \text{Prob}(m^* = M+j) = \prod_{i=1}^{j} \rho_i^* \prod_{i=1}^{14} \rho_i$
j=0	.4354
1	.2215
2	.1268
3	.0769
4	.0482
5	.0310
6	.0203
7	.0135
8	.0090
9	.0061
10	.0042
11	.0028
12	.0020
13	.0014
14	.0009

The figures in Table 2.1 indicate that as $T\to\infty$, the CAT criterion does not prevent asymptotic overfitting of the distributed lag model, i.e. $\lim_{T\to\infty} \text{Prob}(m^*>M) = .5646$. As $T\to\infty$ Parzen's criterion selects the true lag length approximately 44% of the time. The limiting probabilities that $m^* = M+j$ converge rapidly to zero as j increases. Because the use of the CAT criterion results in a non-zero probability that $m^*=M$, there is a gain in asymptotic efficiency when using this criterion to estimate the DLM. Consider the estimator of an unconstrained distributed lag model of unknown order called Hannan inefficient (HI), (7). To get consistent estimates of the coefficients using HI, the fitted distributed lag m(T) must be expanded to infinity with sample size, while the ratio of m(T) over T goes to zero, (17, p. 304). The terminology "Hannan inefficient" stems from the fact that the lag distribution must be expanded indefinitely in <u>both</u> directions, so there is no way of incorporating prior information on the lag distribution (one-sidedness or known lag length) into the estimation procedure. Our analysis is comparable to the HI procedure if model 1.3 is amended to include M future x's. None of the preceding analysis (propositions 1-3) is affected if we fit symmetric <u>two-</u>sided lag distributions of order m(T). Once the random variables $\hat{\sigma}_j^{-2}$, j=1, ..., m(T) are redefined as the error sum of squares from a two sided distributed lag with j leads and lags divided by T-2j-1, the previous results follow with appropriate changes in degrees of freedom. The salient feature of these results is that $\lim_{T\to\infty} \text{Prob}(m^*=M) > 0$. Also, since the limiting probability that $m^* = M+j$ goes to zero as j gets large, for $\delta > 0$ there exists some integer $M_0 = M_0(\delta)$ such that $\lim_{T\to\infty} \text{Prob}(m^* > M_0) < \delta$. Hence there is a gain in asymptotic efficiency over HI when CAT is used to estimate the length of a symmetric two sided distributed lag model. As $T\to\infty$, the CAT criterion selects a <u>finite</u> lag length $m^* \geq M$ with non-zero probability.[14]

We conclude this section with a discussion of the large sample properties of $\hat{\underline{\beta}}_{m^*} = (X'_{m^*}X_{m^*})^{-1}X'_{m^*}y$. We shall focus attention on the first M+1 components of $\hat{\underline{\beta}}_{m^*}$.

Proposition 4: Let $\underline{\hat{\beta}}_{m*}' = (\underline{\hat{\beta}}_M, \hat{\beta}_{m*-M})'$. Then $\underset{T\to\infty}{\text{plim}}\ \underline{\hat{\beta}}_M = \underline{\beta}_M$.

Proof: By proposition 1 we need only consider those values of m* such that $m^* \geq M$. Let $\{x(t)\}$ denote the entire history of the x process, $\{\ldots, x_{t-1}, x_t, x_{t+1}, \ldots\}$. Then

$$\lim_{T\to\infty} E(\hat{\beta}_{m*} \mid \{x(t)\}, m^*) = (X_{m*}'X_{m*})^{-1}X_{m*}'X_M\underline{\beta}_M$$

$$= \begin{pmatrix} \underline{\beta}_M \\ \underline{0} \end{pmatrix} \begin{matrix} (M+1)x1 \\ (m^*-M)x1 \end{matrix} \quad . \tag{1.55}$$

$$\lim_{T\to\infty} \text{Var}(\underline{\hat{\beta}}_{m*} \mid m^*, \{x(t)\}) = \sigma^2(X_{m*}'X_{m*})^{-1}.$$

The (M+1)x(M+1) upper left hand corner block of $\sigma^2(X_{m*}'X_{m*})^{-1}$ is equal to $\sigma^2(X_M'P_{m*-M}X_m)^{-1}$. Denote this matrix by V(m*). The expectation of V(m*) with respect to m* is equal to the variance of the first M+1 components of $\underline{\hat{\beta}}_{m*}$ conditioned on $\{x(t)\}$ alone :

$$E_{m*}(V(m^*) \mid \{x(t)\}, m^*) = \sum_{m^*=0}^{M-1} V(m^*)f(m^*, T \mid \{x(t)\})$$

$$+ \sum_{m^*=M}^{m(T)} V(m^*)f(m^*,T \mid \{x(t)\}), \tag{1.56}$$

where $f(m^*,T|\{x(t)\})$ is the pdf of m* given the x process. The sample size is included as an argument of f(•) to re-emphasize the dependence of m* on T. Since $\underset{T\to\infty}{\text{plim}}\ V(m^*)$ is a well defined matrix (see footnote 15 below), the first term on the right hand side of 1.56 converges to zero as $T\to\infty$ by proposition 1. The second term,

$$\sum_{m^*=M}^{m(T)} V(m^*)f(m^*,T|\{X(t)\}) = \sigma^2 \sum_{m^*=M}^{m(T)} (X_M'P_{m^*-M}X_M)^{-1} f(m^*,T|\{x(t)\})$$

$$\leq \sigma^2 (X_M'P_{m(T)-M}X_M)^{-1} \sum_{m^*=M}^{m(T)} f(m^*,T|\{X(T)\}), \tag{1.57}$$

where $\leq$ denotes matrix ordering. The matrix $P_{m(T)-M}$ is equal to $(I - X_{m(T)-M}(X'_{m(T)-M}X_{m(T)-M})^{-1}X'_{m(T)-M})$. Observe that

$$\operatorname*{plim}_{T\to\infty} \sigma^2 (X_M'P_{m(T)-M}X_M)^{-1} \cdot \sum_{m^*=M}^{m(T)} f(m^*, T|\{x(t)\}) =$$

$$\lim_{T\to\infty} \left(\frac{\sigma^2}{T}\right) \operatorname*{plim}_{T\to\infty} \left(\frac{X_M'P_{m(T)-M}X_M}{T}\right)^{-1} \cdot 1 = 0, \tag{1.58}$$

since $\operatorname*{plim}_{T\to\infty} \left(\frac{X_M'P_{m(T)-M}X_M}{T}\right)^{-1}$ is a well defined matrix.[15]

Therefore, the first M+1 components of $\hat{\underline{\beta}}_{m^*}$ consistently estimate $\underline{\beta}_M$.

Conditional on $\{x(t)\}$ and a <u>fixed</u> $m^* \geq M$, the limiting distribution of $\sqrt{T}(\hat{\underline{\beta}}_{m^*} - \binom{\underline{\beta}_M}{\underline{0}})$ is normal with zero mean vector and covariance matrix $\sigma^2\, Q(m^*)^{-1}$. This is true because for <u>fixed</u> $m^* \geq M$, $\frac{1}{\sqrt{T}}\, X'_{m^*}\, \underline{\varepsilon} \sim N(\underline{0},\ \sigma^2(\frac{X'_{m^*}X_{m^*}}{T}))$ and $\operatorname*{plim}_{T\to\infty} (\frac{X'_{m^*}X_{m^*}}{T})^{-1} = Q(m^*)^{-1}$. The innovation in our analysis has been the estimation of <u>both</u> M and the distributed lag coeff-cients, but it is applied work that motivates this discussion of the conditional limiting distribution of $\hat{\underline{\beta}}_{m^*}$ given m* and $\{x(T)\}$. Once m* has been selected using the CAT criterion, it is convenient for the researcher to act as if m* were fixed when performing hypothesis tests using coefficient estimates and their estimated covariance matrix. In the next section we examine the bias associated with conventional coefficient t-tests when CAT is used to select a DLM from a set of competing models of various lag lengths.

Although we have worked out the limiting probabilities that $m^* = M+i$, $i=0, 1, \ldots, m(T)$ for a distributed lag regression model with an arbitrary zero mean covariance stationary x process and normal independent disturbances, the problem of determining the limiting distribution of $\sqrt{T}(\hat{\underline{\beta}}_{m^*} - \underline{\beta}_{m^*})$, or some other function of $\hat{\underline{\beta}}_{m^*}$, remains a difficult problem. The column dimension of X_{m^*} is a random variable with no upper bound, and "unconventional" central limit theorems are required to examine the limiting distribution of $(1/\sqrt{T})X'_{m^*}\underline{\varepsilon}$. This is true even if one restricts attention to $\hat{\underline{\beta}}_M = (X'_M P_{m^*-M} X_M)^{-1} X'_M P_{m^*-M}\underline{\varepsilon}$, the first M+1 components of $\hat{\underline{\beta}}_{m^*}$. The paper by Sims (17) discusses the problems associated with infinite dimensional parameter spaces; we shall not pursue the subject further.

3. A Monte Carlo study of the use of the CAT criterion to select the order of the distributed lag model.

In this section we report results from a series of Monte Carlo experiments in which the CAT criterion was used to select the length of the coefficient lag distribution in the regression model 1.3 of section one,

$$y(t) = \sum_{s=0}^{M} \beta(s)x(t-s) + \varepsilon(t), \quad t=1, \ldots, T,$$

when M is an unknown nonnegative integer. The explanatory variable x(t) was generated by the covariance stationary process

$$x(t)(1-.8L)^2 = \varepsilon^*(t), \quad t=1, \ldots, T, \tag{1.59}$$

where L denotes the lag operator. Both $\varepsilon^*(t)$ and $\varepsilon(t)$ are "pseudo-random" standard normal variates independent of one another.[16] This particular parameterization for the x(t) process was chosen since the autocovariance function of x(t) closely resembles that of a typical U.S. time series.[17] For each experiment we chose samples of size 50, 100, and 200, and one of the following lag distributions:

a) No lag distribution; $\beta(0) = 1.0$ and $\beta(i) = 0$, $i \neq 0$.

b) Linear Decay; $\beta(i) = 1.0$, $0 \leq i \leq 4$ and $\beta(i) = 1.0 - .1(i-4)$, $5 \leq i \leq 13$. (1.60)

c) Box; $\beta(0) = .5 = \beta(6)$, $\beta(i) = .8$, $i \leq i \leq 5$.

d) Infinite geometric; $\beta(i) = .8^i$, $i = 0, 1, \ldots$.[18]

For each sample size the maximum order of the fitted distributed lag models was approximately $T^{.6}$; when T=50, distributed lags of order 0-12 were fit, for T=100, 0-18, and for T=200, 0-27. For each replication and sample size the CAT criterion estimated a particular lag distribution. After 100 replications the following summary statistics (i-viii) were calculated.

(i) The average order of the lag distribution that was selected by minimixing CAT(i), i=0, 1, ..., $(T^{.6})$, denoted m*(T). Let m*(T,k) denote the order chosen for sample size T and replication k. The average order m*(T) is given by

$$m^*(T) = \frac{1}{100} \sum_{k=1}^{100} m^*(T,k), \quad T=50,\ 100,\ 200. \qquad (1.61)$$

We include this statistic in the analysis since it gives a general indication of the performance of CAT as sample size increases. For example, when the CAT criterion is applied to the finite lag distributions (a), (b), (c), we would expect the average m* to more closely coincide with the population lag length M, the larger is the sample size. When CAT is used to estimate the lag distribution which is of infinite length, lag distribution (d), we would expect the average m*(T) to be an increasing function of sample size.

(ii) The average estimate of σ^2, the variance of the disturbance term (σ^2 = Var ($\varepsilon(t)$) = 1), denoted $\hat{\sigma}^2(T)$. Let $\hat{\sigma}^2(T,k)$ denote the estimate of σ^2 for sample size T and replication k. Then $\hat{\sigma}^2(T)$ is given by

$$\hat{\sigma}^2(T) = \frac{1}{100} \sum_{k=1}^{100} \hat{\sigma}^2(T,k), \quad T=50,\ 100,\ 200. \qquad (1.62)$$

This statistic is important since the residual variance is one component of the estimated variance of the coefficient estimates. If $\hat{\sigma}^2$ is biased downwards (upwards), we would expect coefficient t statistics to be too large (small) on average, provided the coefficient estimator is unbiased. In any event we would expect bias in $\hat{\sigma}^2(T)$ to diminish as sample size increases.

(iii) A Komolgorov-Smirnov (K-S) statistic to test the null hypothesis that the sum of the estimated coefficients was equal to the sum of the true coefficients. Let CSUM(T,k) denote the sum of the coefficients from the estimated lag distribution of order m*(T,k) minus the sum of the population distributed lag coefficients, divided by the $(T,k)^{th}$ standard error of the estimated sum. Assume that the CSUM(T,k) are independent and identically distributed normal variates with zero mean and unit variance, denoted N(0,1)[19] Suppose CSUM(T,k) is re-indexed by CSUM(T,ℓ) so that the CSUM(T,ℓ), ℓ=1, ... 100 form a monotone increasing sequence. Let Φ(CSUM(T,ℓ)) denote the cumulative density function (CDF) of a N(0,1) evaluated at CSUM(T,ℓ). The two-sided K-S(T), T=50, 100, 200 statistics reported below are equal to the maximum absolute difference between the sample and population CDF of CSUM(T,ℓ),

$$K\text{-}S(T) = \max_{\ell=1,\,\ldots,\,100} \left| \frac{\ell}{100} - \Phi(\mathrm{CSUM}(T,\ell)) \right|, \qquad T=50,\ 100,\ 200. \tag{1.63}$$

The null hypothesis that the estimated sum is equal to the true sum is rejected at the 5% (1%) significance level if the K-S(T) statistic exceeds .136(.163). An asterisk denotes acceptance of the null hypothesis at the 5% significance level.

We choose to report a K-S statistic for the sum of the estimated coefficients since for some distributed lag models this sum represents the cumulative or long run response of an endogenous variable to a once and for all change in an exogenous variable. The researcher may want to know if the use of the CAT criterion distorts this statistic. It should be noted that the same information concerning the distribution of the sum of the estimated coefficients could have been obtained from a standard t-statistic. Failure to do so was an oversight on the author's part.

(iv) The average bias of the coefficient estimates, denoted BIAS(i,T). Let $\beta(i,T,k)$ denote the ith estimated coefficient i=0, 1, ..., m*(T,k) for sample size T and replication k, and let $\beta(i)$ denote the value of the ith population coefficient from 2.12 a-d. For T=50, 100, 200 define $\beta^*(i,T,k)$ as

$$\hat{\beta}^*(i,T,k) = \begin{cases} \hat{\beta}(i,T,k) \text{ if } i=0,\ 1,\ \ldots,\ m^*(T,k), \text{ and} \\ 0 \text{ if } i = m^*(T,k) + 1,\ \ldots,\ T^{.6}. \end{cases} \tag{1.64}$$

Then BIAS (i,T) is given by

$$\text{BIAS}(i,T) = \frac{1}{100} \sum_{k=1}^{100} \hat{\beta}^*(i,T,k) - \beta(i), \tag{1.65}$$

$$i=0,\ 1,\ \ldots,\ T^{.6}, \text{ and } T=50,\ 100,\ 200.$$

This statistic is important since it helps determine the reliability of coefficient point estimates when the lag distribution has been estimated using the CAT criterion.

(v) The number of times that coefficient (i,T) was included in the estimated lag distribution, denoted NTIME(i,T). Define the variable Q(i,T,k) as follows;

$$Q(i,T,k) = \begin{cases} 1 \text{ if } i \leq m^*(T,k), \text{ and} \\ 0 \text{ if } i = m^*(T,k) + 1,\ \ldots,\ T^{.6}. \end{cases} \tag{1.66}$$

Then NTIME(i,T) is given by

$$\text{NTIME}(i,T) = \sum_{k=1}^{100} Q(i,T,k), \quad i=0,\ 1,\ \ldots,\ T^{.6}, \qquad (1.67)$$

$$T=50,\ 100,\ 200.$$

The values NTIME(i,T) $i=0, 1, \ldots, T^{.6}$, $T=50, 100, 200$ can be used to compute the sample frequencies that $m^*(T) = i$, $i=0, \ldots, (T^{.6})$ which can then be compared to the limiting probabilities that $m^* = i$, $i=0, \ldots, (M+14)$ that were calculated in section two pages 18-26. The sample frequency that $m^*(T) = i$ for any sample size and lag distribution is given by

$$(\text{NTIME}(i,T) - \text{NTIME}(i+1,T))/100 \text{ for } i=0, \ldots, (T^{.6}-1), \qquad (1.68)$$

and $\text{NTIME}(i,T)/100$ for $i=T^{.6}$.

We expect there to be greater coincidence between the sample and limiting probabilities that $m^*(T) = i$, $i=0, 1, \ldots, T^{.6}$, the larger is the sample size.

(vi) The average mean square error of the estimated coefficients, denoted MSE(i,T)

$$\text{MSE}(i,T) = \frac{1}{100} \sum_{k=1}^{100} (\hat{\beta}^*(i,T,k) - \beta(i))^2, \qquad (1.69)$$

$$i=0,\ 1,\ \ldots,\ T^{.6},\ T=50,\ 100,\ 200.$$

This statistic is useful since we will compare it to the average estimated variance of each coefficient EVAR(i,T), $i=0, \ldots, T^{.6}$, $T=50, 100, 200$, which is described below.

(vii) The average estimated variance of coefficient (i,T), denoted EVAR(i,T). Let $\hat{\sigma}^2(T,k)\cdot(X(T,k)'\cdot X(T,k))^{-1}_{ii}$ denote the diagonal elements of the estimated covariance matrix of the $\hat{\beta}(i,T,k)$, $i=0, 1, \ldots, T^{.6}$. Define

$$\hat{\sigma}^{*2}(T,k)(X^*(T,k)' \cdot X^*(T,k))^{-1}_{ii} = \begin{cases} \hat{\sigma}^2(T,k)(X(T,k)' \cdot X(T,k))^{-1}_{ii} \\ \text{for } i=0,1,\ \ldots,\ m^*(T,k), \text{ and} \\ 0 \text{ for } i=m^*(T,k)+1,\ \ldots,\ T^{.6}. \end{cases} \tag{1.70}$$

Then EVAR(i,T) is given by

$$\text{EVAR}(i,T) = \frac{1}{100} \sum_{k=1}^{100} \hat{\sigma}^{*2}(X^*(T,k)' \cdot X^*(T,k))^{-1}_{ii},$$

$$i=1,\ \ldots,\ T^{.6},\ T=50,\ 100,\ 200.$$

If BIAS(i,T) is small, we expect EVAR(i,T) and MSE(i,T) to be roughly the same. Should EVAR(i,T) be biased upwards (downwards) then coefficient t-tests will be too small (large). We examine the bias of coefficient hypothesis tests under the null hypothesis that $\beta(i) = 0$, $i=0, 1, \ldots, T^{.6}$ using F(i,T,k) described below.

(viii) The sample CDF of the ratio of each coefficient to its estimated standard error, assuming the ratio has a N(0,1) distribution. Define F(i,T,k) as

$$F(i,T,k) = \begin{cases} \Phi[\hat{\beta}(i,T,k)/(\hat{\sigma}^2(T,k) \cdot (X(T,k)'X(T,k))^{-1}_{ii})^{1/2}] \text{ for} \\ i=0,\ 1,\ \ldots,\ m^*(T,k), \text{ and} \\ .5 \text{ for } i=m^*(T,k)+1,\ \ldots,\ T^{.6}. \end{cases} \tag{1.71}$$

If the CAT criterion selected an estimated lag distribution of order $m^*(T,k) < T^{.6}$, then the $(m^*(T,k) + 1)$ through $(T^{.6})$ -th coefficients were arbitrarily assigned a cumulative probability of .5. The cumulative normal probabilities F(i,T,k) are found in 13 columns of tables 2.3, 2.5, 2.7, and 2.9, a table for each lag distribution (1.60)a-d. For example, the entry corresponding to the column headed by 3 and the row associated with Lag 6 in Table 2.3-Lag Distribution (a) - T=50, page 43 is 2. This means that F(6,50,k) is in the third probability cell; F(6,50,k) is less than .05 and greater than or equal to .025 for 2 out of the 100 replications for lag distribution (a), sample size 50, and the coefficient corresponding to the x variable lagged six periods. The null hypothesis that the ratio of an estimated coefficient to its estimated standard error is distributed as a N(0,1) is incorrect for the first M+1 coefficients of each lag distribution and correct for all others.[20] Theoretically, the sample ratios of those coefficients whose population value is zero should be distributed across the columns of tables 2.3, 2.5, 2.7, and 2.9 in proportion to the cumulative probabilities at the head of each column. For example, given any lag distribution there should be approximately 30 replications in the first five columns in all rows where the population coefficient is zero, since the first five columns represent a cumulative probability of .30. We shall return to this point during the analysis of the Monte Carlo results.

Last, the statistics described in (i-vii) above are found in tables 2.2, 2.4, 2.6, and 2.8; these tables correspond to the four lag distributions 1.60 a-d. The first three statistics (i-iii) are found under the heading general statistics, while (iv-vii) are found under the heading coefficient statistics.

Table 2.2

Lag Distribution (a)

T=50

General Statistics

$m^*(50) = 3.22$ $\hat{\sigma}^2(50) = .8836$ K-S(50) = .0974*

Coefficients Statistics

BIAS(50,i)	NTIME(50,i)	MSE(50,i)	EVAR(50,i)	LAG-i
.0281	100	.0264	.0116	0
-.0040	51	.0755	.0332	1
.0370	45	.0634	.0316	2
-.00898	38	.0701	.0282	3
-.0207	34	.0412	.0251	4
.0135	30	.0439	.0211	5
.0105	25	.0394	.0194	6
-.0135	24	.0378	.0181	7
.00456	21	.0590	.0162	8
-.00225	18	.0466	.0140	9
.00461	16	.0354	.0115	10
-.000409	13	.0264	.00640	11
-.000504	7	.00856	.00152	12

Table 2.2 cont.

Lag Distribution (a)

T=100

General Statistics

m*(100) = 3.07 $\hat{\sigma}^2(100)$ = .9476   K-S(100) = .1013*

Coefficients Statistics

BIAS(100),i)	NTIME(100,i)	MSE(100,i)	EVAR(100,i)	LAG-i
.0178	100	.00902	.00543	0
-.0213	56	.0254	.0157	1
.0122	41	.0247	.0101	2
.0104	34	.0273	.0120	3
-.0274	28	.0199	.0107	4
.0177	25	.0169	.00949	5
.00356	21	.0180	.00845	6
-.0211	18	.0154	.00760	7
.0136	15	.0212	.00694	8
-.0000761	13	.0146	.00627	9
-.00212	12	.00755	.00485	10
.000331	8	.00678	.00433	11
.0111	8	.00570	.00431	12
-.0121	8	.00753	.00399	13
-.00256	7	.00921	.00304	14
.0114	5	.00984	.00223	15
-.00753	4	.00440	.00163	16
.00233	3	.000761	.000705	17
.000430	1	.0000185	.000110	18

Table 2.2 cont.

Lag Distribution (a)

T=200

General Statistics

m*(200) = 3.11 $\hat{\sigma}^2(200)$ = .9747 K-S(200) = .0778*

Coefficients Statistics

BIAS(200,i)	NTIME(200,i)	MSE(200,i)	EVAR(200,i)	LAG-i
.00856	100	.00372	.00230	0
-.00993	51	.0134	.00656	1
.00718	36	.0142	.00632	2
.00480	32	.0116	.00524	3
-.0137	26	.00952	.00425	4
.000530	21	.00623	.00356	5
.00488	17	.00641	.00317	6
.00207	16	.00433	.00264	7
-.00140	12	.00426	.00233	8
.00399	11	.00124	.00218	9
-.00667	10	.00239	.00212	10
.00528	10	.00331	.00202	11
.000179	9	.00259	.00196	12
-.00381	9	.00290	.00182	13
.0116	8	.00629	.00168	14
-.00910	8	.00289	.00128	15
.000691	5	.000579	.00111	16
-.00411	5	.000949	.00109	17
.00780	5	.00253	.000958	18
-.00371	4	.00118	.000913	19
.00430	4	.00174	.000860	20
-.00364	4	.00104	.000621	21
-.00147	3	.000127	.000402	22
.00101	2	.000186	.000330	23
.000744	2	.0000993	.000210	24
.0000909	1	.000000826	.000504	25
-	0	-	-	26
-	0	-	-	27

Table 2.3 - Lag Distribution (a)

T=50

Number of replications in each probability cell

	1	2	3	4	5	6	7	8	9	10	11	12	13
	Probabilities												
Lag	0.0-.0125	.0125-.025	.025-.05	.05-.1	.1-.3	.3-.5	.5	.5-.7	.7-.9	.9-.95	.95-.975	.975-.9875	.9875-1.0
0	0	0	0	0	0	0	0	0	0	0	0	0	100
1	6	3	2	6	11	1	49	5	1	1	3	0	2
2	3	1	1	2	7	5	55	6	5	7	0	4	4
3	6	0	1	2	7	7	62	2	4	2	1	1	5
4	1	3	2	2	6	6	66	4	4	2	2	2	0
5	2	2	1	0	1	5	70	5	5	3	3	2	1
6	4	0	2	0	2	1	75	3	8	2	0	3	0
7	3	2	1	0	4	2	76	6	4	0	1	0	1
8	4	1	0	0	2	2	79	1	5	1	0	1	4
9	2	0	0	5	1	0	82	3	4	0	0	2	1
10	2	0	1	2	2	2	84	0	1	0	1	3	2
11	4	1	1	0	1	0	87	0	1	2	1	1	1
12	2	0	1	0	0	0	93	1	0	0	0	1	2

T=100

0	0	0	0	0	0	0	0	0	0	0	0	0	100
1	4	3	5	5	6	11	44	4	8	3	4	1	2
2	2	0	1	4	5	5	59	9	7	1	2	1	4
3	4	1	1	2	5	3	66	3	7	3	3	0	2
4	4	1	1	2	7	2	72	3	4	2	1	0	1
5	1	0	0	2	4	2	75	4	5	2	1	1	3

Table 2.3 - Lag Distribution (a)

T=100

Number of replications in each probability cell

	1	2	3	4	5	6	7	8	9	10	11	12	13
	Probabilities												
Lag	0.0-.0125	.0125-.025	.025-.05	.05-.1	.1-.3	.3-.5	.5	.5-.7	.7-.9	.9-.95	.95-.975	.975-.9875	.9875-1.0
6	1	2	1	1	5	0	79	4	1	1	1	1	3
7	2	1	1	0	7	1	82	2	3	0	0	0	1
8	0	1	1	1	2	1	85	2	2	1	1	0	3
9	2	1	0	1	0	1	87	3	3	1	0	0	1
10	1	0	1	0	2	1	88	3	2	0	0	0	2
11	0	0	1	1	1	1	92	1	1	1	1	0	0
12	0	0	0	0	1	0	92	3	3	0	0	1	0
13	0	1	1	2	0	0	92	3	0	1	0	0	0
14	1	0	0	0	2	0	93	1	0	2	0	0	1
15	0	0	0	1	0	1	95	0	2	0	0	0	1
16	1	0	1	0	1	0	96	0	1	0	0	0	0
17	0	0	0	0	1	0	97	0	0	2	0	0	0
18	0	0	0	0	0	0	99	1	0	0	0	0	0
							T=200						
0	0	0	0	0	0	0	0	0	0	0	0	0	100
1	4	4	4	4	8	6	49	5	7	3	2	2	2
2	2	1	1	2	6	3	64	5	6	3	2	1	4
3	1	3	3	1	2	6	68	4	4	3	0	1	4
4	5	0	2	3	3	4	74	1	4	1	3	0	0
5	0	0	2	3	4	2	79	3	1	0	3	1	2
6	1	1	0	1	2	2	83	3	3	0	1	1	2

Table 2.3 - Lag Distribution (a)

T= 200

Number of replications in each probability cell

Probabilities

	1	2	3	4	5	6	7	8	9	10	11	12	13
Lag	0.0- .0125	.0125- .025	.025- .05	.05- .1	.1- .3	.3- .5	.5	.5- .7	.7- .9	.9- .95	.95- .975	.975- .9875	.9875- 1.0
7	1	1	0	1	1	3	84	4	2	0	1	0	2
8	1	0	0	1	3	2	88	0	3	1	1	0	0
9	0	0	0	1	1	4	89	2	2	1	0	0	0
10	0	1	1	1	1	2	90	2	2	0	0	0	0
11	0	0	0	1	2	1	90	1	4	0	0	0	1
12	0	0	0	2	2	0	91	1	2	2	0	0	0
13	1	0	0	0	3	1	91	0	4	0	0	0	0
14	0	0	0	1	0	3	92	0	0	0	0	1	3
15	1	2	1	0	0	1	92	1	2	0	0	0	0
16	0	0	0	0	1	1	95	1	2	0	0	0	0
17	0	0	0	1	2	0	95	1	1	0	0	0	0
18	0	0	0	0	0	1	95	1	1	0	0	2	0
19	0	0	0	1	1	1	96	0	1	0	0	0	0
20	0	0	0	0	1	0	96	1	0	1	1	0	0
21	1	0	0	0	2	0	96	0	1	0	0	0	0
22	0	0	0	0	1	2	97	0	0	0	0	0	0
23	0	0	0	0	0	1	98	0	1	0	0	0	0
24	0	0	0	0	0	1	98	0	1	0	0	0	0
25	0	0	0	0	0	1	99	0	0	0	0	0	0
26	0	0	0	0	0	0	100	0	0	0	0	0	0
27	0	0	0	0	0	0	100	0	0	0	0	0	0

Table 2.4

Lag Distribution (b)

T=50

General Statistics

m*(50) = 11.17 $\hat{\sigma}^2(50)$ = .9678 K-S(50) = .0645*

Coefficient Statistics

BIAS(50,i)	NTIME(50,i)	MSE(50,i)	EVAR(50,i)	LAG-i
-.000726	100	.0413	.0386	0
.00821	100	.143	.129	1
-.00368	100	.169	.144	2
.0289	100	.126	.142	3
-.0309	100	.101	.142	4
-.00246	100	.141	.141	5
.00548	100	.150	.140	6
-.0107	100	.152	.142	7
.0255	100	.194	.143	8
-.0403	100	.224	.137	9
.0436	98	.193	.109	10
-.0143	74	.139	.0655	11
-.0151	45	.0394	.0170	12

Table 2.4 cont.

Lag Distribution (b)

T=100

General Statistics

$m^*(100) = 12.65 \quad \hat{\sigma}^2(100) = .9709 \quad \text{K-S}(100) = .0705^*$

Coefficient Statistics

BIAS(100,i)	NTIME(100,i)	MSE(100,i)	EVAR(100,i)	LAG-i
-.0124	100	.0125	.0134	0
.03.2	100	.0507	.0472	1
-.0367	100	.0431	.0531	2
.0321	100	.0483	.0533	3
-.0130	100	.0659	.0533	4
-.000332	100	.0626	.0535	5
-.0162	100	.0655	.0533	6
.0141	100	.0575	.0533	7
.00265	100	.0550	.0534	8
.0164	100	.0636	.0535	9
-.0225	100	.0507	.0509	10
.00419	99	.0632	.0366	11
-.0150	63	.0506	.0229	12
.0265	39	.0312	.0150	13
-.00234	28	.0197	.00899	14
-.0134	16	.0100	.00532	15
-.00139	10	.00757	.00335	16
.00635	6	.00759	.00196	17
-.00189	4	.00119	.000479	18

Table 2.4 cont.

Lag Distribution (b)

T=200

General Statistics

m*(200) = 13.60 $\hat{\sigma}^2(200)$ = .9726 K-S(200) = .0917*

Coefficient Statistics

BIAS(200,i)	NTIME(200,i)	MSE(200,i)	EVAR(200,i)	LAG-i
-.00547	100	.00613	.00578	0
.0158	100	.0178	.0205	1
-.0287	100	.0180	.0229	2
.0323	100	.0255	.0229	3
-.0205	100	.0246	.0229	4
-.000843	100	.0220	.0228	5
.00678	100	.0242	.0228	6
.0681	100	.0207	.0228	7
-.0135	100	.0201	.0229	8
.00826	100	.0251	.0230	9
-.000897	100	.0242	.0225	10
-.00953	100	.0344	.0185	11
.00990	78	.0234	.0121	12
.00232	47	.0128	.00762	13
-.00654	31	.0178	.00522	14
-.000163	23	.0117	.00393	15
-.000289	18	.00420	.00304	16
.00124	14	.00436	.00240	17
-.00226	11	.00507	.00201	18
.00577	10	.00408	.00153	19
-.00342	7	.00378	.00120	20
.000717	6	.00309	.00100	21
-.000182	5	.00150	.000814	22
.000406	4	.00116	.000602	23
.000904	3	.00108	.000424	24
-.00106	2	.000474	.000230	25
-.00128	1	.000164	.000230	26
-	0	-	-	27

Table 2.5 - Lag Distribution (b)

T=50

Number of replications in each probability cell

	1	2	3	4	5	6	7	8	9	10	11	12	13
	Probabilities												
Lag	0.0-.0125	.0125-.025	.025-.05	.05-.1	.1-.3	.3-.5	.5	.5-.7	.7-.9	.9-.95	.95-.975	.975-.9875	.9875-1.0
0	0	0	0	0	0	0	0	0	0	0	0	0	100
1	0	0	0	0	0	1	0	0	8	6	5	6	74
2	0	0	0	0	0	0	0	2	12	4	7	9	66
3	0	0	0	0	0	1	0	0	8	6	6	2	77
4	0	0	0	0	0	0	0	3	6	13	12	11	55
5	0	0	0	0	0	1	0	3	16	11	9	8	52
6	0	0	0	0	1	2	0	3	19	11	11	15	38
7	0	0	0	0	3	5	0	6	18	18	10	12	28
8	0	0	1	0	5	6	0	10	17	14	11	7	29
9	0	1	0	3	6	14	0	12	21	10	10	5	18
10	0	1	0	3	4	8	2	19	22	5	3	5	28
11	0	0	0	5	10	7	26	5	11	7	5	3	21
12	0	1	2	0	3	6	55	3	9	4	6	1	10
							T=100						
0	0	0	0	0	0	0	0	0	0	0	0	0	100
1	0	0	0	0	0	0	0	0	0	0	0	1	99
2	0	0	0	0	0	0	0	0	0	1	1	0	98
3	0	0	0	0	0	0	0	0	0	0	0	0	100
4	0	0	0	0	0	0	0	0	0	1	4	3	92
5	0	0	0	0	0	0	0	0	4	1	4	2	89

Table 2.5 - Lag Distribution (b)

T= 100

Number of replications in each probability cell

	1	2	3	4	5	6	7	8	9	10	11	12	13
	Probabilities												
Lag	0.0-.0125	.0125-.025	.025-.05	.05-.1	.1-.3	.3-.5	.5	.5-.7	.7-.9	.9-.95	.95-.975	.975-.9875	.9875-1.0
6	0	0	0	0	0	1	0	0	3	5	12	10	69
7	0	0	0	0	0	2	0	1	8	6	6	9	68
8	0	0	0	0	0	2	0	3	16	8	8	14	49
9	0	0	0	1	0	2	0	9	21	9	8	13	37
10	0	0	0	0	2	7	0	16	29	14	6	4	22
11	0	0	1	1	11	7	1	13	13	4	8	7	24
12	1	1	1	3	6	2	37	7	12	5	7	7	11
13	3	0	1	2	4	3	61	8	6	4	4	2	2
14	2	1	1	3	5	5	72	3	2	2	1	0	3
15	1	2	0	2	2	3	84	2	2	1	0	1	0
16	2	0	0	0	1	2	90	1	1	0	1	0	2
17	0	0	0	1	0	2	94	0	2	0	0	0	1
18	0	1	0	0	2	0	96	0	0	0	1	0	0
	T=200												
0	0	0	0	0	0	0	0	0	0	0	0	0	100
1	0	0	0	0	0	0	0	0	0	0	0	0	100
2	0	0	0	0	0	0	0	0	0	0	0	0	100
3	0	0	0	0	0	0	0	0	0	0	0	0	100
4	0	0	0	0	0	0	0	0	0	0	0	0	100
5	0	0	0	0	0	0	0	0	0	0	0	1	99
6	0	0	0	0	0	0	0	0	0	0	0	0	100

Table 2.5 - Lag Distribution (b)

T=200

Number of replications in each probability cell

	1	2	3	4	5	6	7	8	9	10	11	12	13
	Probabilities												
Lag	0.0-.0125	.0125-.025	.025-.05	.05-.1	.1-.3	.3-.5	.5	.5-.7	.7-.9	.9-.95	.95-.975	.975-.9875	.9875-1.0
7	0	0	0	0	0	0	0	0	0	0	3	1	96
8	0	0	0	0	0	0	0	0	3	3	4	6	84
9	0	0	0	0	0	0	0	1	9	8	6	8	68
10	0	0	0	0	1	1	0	2	21	15	10	12	38
11	0	0	0	2	6	7	0	11	23	12	3	2	34
12	0	0	1	1	6	5	22	9	12	9	6	7	22
13	1	2	3	5	6	7	53	2	11	3	3	3	1
14	3	2	4	5	2	2	69	3	3	1	1	1	4
15	3	0	2	1	3	2	77	1	4	2	2	1	2
16	0	0	0	3	3	4	82	2	2	1	1	0	2
17	1	0	0	2	2	2	86	1	3	1	1	0	1
18	0	0	0	2	5	0	89	0	3	0	0	0	1
19	1	0	0	1	1	0	90	4	0	0	0	2	1
20	2	0	0	0	0	2	93	0	2	1	0	0	0
21	1	0	1	0	0	0	94	2	0	0	0	0	2
22	0	0	0	0	3	0	95	1	0	0	0	0	1
23	0	0	1	0	0	1	96	1	0	0	1	0	0
24	1	0	0	0	0	1	97	0	0	0	0	1	0
25	1	0	0	0	0	0	98	0	1	0	0	0	0
26	0	0	1	0	0	0	99	0	0	0	0	0	0
27	0	0	0	0	0	0	100	0	0	0	0	0	0

Table 2.6

Lag Distribution (c)

T=50

General Statistics

m*(50) = 7.52 $\hat{\sigma}^2(50)$ = .9232 K-S(50) = .0752*

Coefficient Statistics

BIAS(50,i)	NTIME(50,i)	MSE(50,i)	EVAR(50,i)	LAG-i
-.0237	100	.0357	.0290	0
.0273	100	.116	.0951	1
.0133	100	.127	.106	2
-.0319	100	.0839	.106	3
.0445	100	.0936	.106	4
-.0181	100	.138	.0966	5
-.0541	96	.121	.0635	6
.0238	51	.0999	.0393	7
.00714	36	.0580	.0281	8
.0149	27	.0586	.0198	9
-.0150	19	.0398	.0143	10
-.00698	15	.0324	.00928	11
.00625	8	.0102	.00253	12

Table 2.6 cont.

Lag Distribution (c)

T=100

General Statistics

$m^*(100) = 8.18 \quad \hat{\sigma}^2(100) = .9665 \quad \text{K-S}(100) = .0581^*$

Coefficient Statistics

BIAS(100,i)	NTIME(100,i)	MSE(100,i)	EVAR(100,i)	LAG-i
-.0206	100	.0127	.0119	0
.0363	100	.0458	.0415	1
.00869	100	.0535	.0466	2
-.0246	100	.0431	.0465	3
-.0137	100	.0476	.0465	4
.0279	100	.0517	.0439	5
-.0328	100	.0332	.0269	6
.0305	46	.0332	.0159	7
-.0204	32	.0208	.0122	8
.0157	26	.0158	.0100	9
-.0101	22	.0232	.00911	10
.0100	21	.0238	.00798	11
-.00350	17	.0147	.00737	12
-.00918	17	.0117	.00628	13
.0149	13	.0132	.00518	14
-.0117	11	.00826	.00376	15
.00518	7	.00595	.00201	16
.000177	4	.00286	.000941	17
-.000429	2	.00135	.000225	18

Table 2.6 cont.

Lag Distribution (c)

T=200

General Statistics

m*(200) = 7.86 $\hat{\sigma}^2(200)$ = .9810 K-S(200) = .0801*

Coefficient Statistics

BIAS(200,i)	NTIME(200,i)	MSE(200,i)	EVAR(200,i)	LAG-i
-.0104	100	.00516	.00540	0
.0143	100	.0211	.0189	1
.00793	100	.0282	.0210	2
-.0124	100	.0220	.0211	3
-.00273	100	.0191	.0121	4
.0164	100	.0203	.0200	5
-.0227	100	.0170	.0124	6
.00956	46	.0132	.00712	7
.00296	30	.00592	.00479	8
.00129	20	.00469	.00366	9
-.000682	16	.00858	.00305	10
-.0118	14	.00741	.00245	11
.0111	11	.00507	.00200	12
-.00553	9	.00399	.00173	13
.00643	8	.00393	.00150	14
-.00628	7	.00225	.00109	15
.00405	4	.00113	.000947	16
-.00392	4	.000485	.000923	17
.00138	4	.000699	.000791	18
.00339	3	.000685	.000716	19
-.00392	3	.00134	.000548	20
.00292	2	.00173	.000323	21
-.00216	1	.000466	.000260	22
.000766	1	.0000586	.000257	23
-.000612	1	.0000374	.000257	24
.000992	1	.0000984	.000228	25
-.000362	1	.0000131	.0000648	26
-	0	-	-	27

Table 2.7 - Lag Distribution (c)

T= 50

Number of replications in each probability cell

	1	2	3	4	5	6	7	8	9	10	11	12	13
	Probabilities												
Lag	0.0-.0125	.0125-.025	.025-.05	.05-.1	.1-.3	.3-.5	.5	.5-.7	.7-.9	.9-.95	.95-.975	.975-.9875	.9875-1.0
0	0	0	0	0	0	1	0	1	6	7	7	7	71
1	0	0	0	0	0	0	0	2	10	8	7	5	68
2	0	0	0	0	1	0	0	2	10	7	5	12	63
3	0	0	0	0	0	0	0	2	8	10	12	6	63
4	0	0	0	0	0	1	0	1	4	8	6	11	69
5	0	0	0	0	1	1	0	3	12	7	5	8	63
6	0	0	0	1	5	4	4	6	7	7	5	12	49
7	5	0	4	3	5	6	49	6	6	5	3	2	6
8	3	0	2	3	5	4	64	3	8	3	1	0	4
9	2	3	0	1	4	1	73	5	4	1	0	2	4
10	3	0	1	2	4	0	81	3	3	0	0	0	3
11	2	2	0	1	1	0	85	3	2	1	2	0	1
12	0	1	1	0	1	1	92	0	1	1	1	0	1

T= 100

	1	2	3	4	5	6	7	8	9	10	11	12	13
0	0	0	0	0	0	0	0	0	1	1	1	0	97
1	0	0	0	0	0	0	0	0	0	3	2	1	94
2	0	0	0	0	0	0	0	0	1	3	2	2	92
3	0	0	0	0	0	0	0	0	0	1	4	2	93
4	0	0	0	0	0	0	0	0	0	1	1	7	91
5	0	0	0	0	0	0	0	0	1	1	2	4	92

Table 2.7 - Lag Distribution (c)

T=100

Number of replications in each probability cell

	1	2	3	4	5	6	7	8	9	10	11	12	13
	Probabilities												
Lag	0.0-.0125	.0125-.025	.025-.05	.05-.1	.1-.3	.3-.5	.5	.5-.7	.7-.9	.9-.95	.95-.975	.975-.9875	.9875-1.0
6	0	0	0	0	0	0	0	2	10	3	8	1	76
7	1	1	4	2	3	3	54	10	9	4	3	3	3
8	3	3	1	2	6	1	68	6	8	0	2	0	0
9	0	0	1	1	7	1	74	5	5	1	2	2	1
10	1	2	3	1	2	1	78	4	3	1	2	0	2
11	2	1	0	1	4	2	79	2	3	1	1	2	2
12	1	1	2	0	2	3	83	2	2	1	1	1	1
13	2	0	1	1	4	1	83	2	1	1	3	0	1
14	1	0	0	1	2	1	87	0	3	3	0	0	2
15	2	1	0	1	1	1	89	3	0	1	0	1	0
16	0	0	1	0	0	1	93	2	2	0	0	0	1
17	0	0	1	1	1	0	96	0	0	0	0	1	0
18	1	0	0	0	0	0	98	0	0	0	0	0	1
							T=200						
0	0	0	0	0	0	0	0	0	0	0	0	0	100
1	0	0	0	0	0	0	0	0	0	0	0	0	100
2	0	0	0	0	0	0	0	0	0	0	0	0	100
3	0	0	0	0	0	0	0	0	0	0	0	0	100
4	0	0	0	0	0	0	0	0	0	0	0	0	100
5	0	0	0	0	0	0	0	0	0	0	0	0	100
6	0	0	0	0	0	0	0	0	1	5	3	2	89

Table 2.7 - Lag Distribution (c)

T= 200

Number of replications in each probability cell

	1	2	3	4	5	6	7	8	9	10	11	12	13
	Probabilities												
Lag	0.0-.0125	.0125-.025	.025-.05	.05-.1	.1-.3	.3-.5	.5	.5-.7	.7-.9	.9-.95	.95-.975	.975-.9875	.9875-1.0
7	2	2	3	1	5	10	54	5	4	3	6	1	4
8	3	1	1	0	5	3	70	2	12	1	1	1	0
9	0	1	1	2	4	1	80	5	3	1	1	0	1
10	1	2	1	1	2	1	84	1	3	0	2	2	0
11	3	0	1	2	1	2	86	1	1	1	1	0	1
12	0	1	0	0	2	0	89	0	4	1	1	0	2
13	1	0	1	1	2	1	91	0	1	1	0	0	1
14	0	0	1	1	0	1	92	0	1	2	0	1	1
15	1	2	0	0	1	0	93	1	1	1	0	0	0
16	0	0	0	0	0	2	96	0	0	2	0	0	0
17	0	0	0	0	2	2	96	0	0	0	0	0	0
18	0	0	0	1	0	0	96	2	0	0	1	0	0
19	0	0	0	0	0	0	97	2	0	0	1	0	0
20	0	1	0	1	0	0	97	1	0	0	0	0	0
21	0	0	0	1	0	0	98	0	0	0	0	0	1
22	0	0	0	1	0	0	99	0	0	0	0	0	0
23	0	0	0	0	0	0	99	1	0	0	0	0	0
24	0	0	0	0	0	1	99	0	0	0	0	0	0
25	0	0	0	0	0	0	99	0	1	0	0	0	0
26	0	0	0	0	0	1	99	0	0	0	0	0	0
27	0	0	0	0	0	0	100	0	0	0	0	0	0

Table 2.8

Lag Distribution (d)

T=50

General Statistics

$m^*(50) = 11.53$ $\hat{\sigma}^2(50) = 1.280$ K-S(50) = .7004

Coefficient Statistics

BIAS(50,i)	NTIME(50,i)	MSE(50,i)	EVAR(50,i)	LAG-i
-.110	100	.0555	.0518	0
.0718	100	.114	.171	1
.0624	100	.139	.193	2
-.0782	100	.144	.195	3
.0232	100	.157	.194	4
-.0312	100	.163	.193	5
.0386	100	.171	.191	6
-.0336	100	.202	.191	7
.0125	100	.189	.187	8
-.0611	99	.166	.183	9
.0380	98	.166	.171	10
-.128	88	.200	.129	11
.267	68	.140	.0344	12

Table 2.8 cont.

Lag Distribution (d)

T=100

General Statistics

$m^*(100) = 16.43$ $\hat{\sigma}^2(100) = 1.024$ K-S(100) = .4030

Coefficient Statistics

BIAS(100,i)	NTIME(100,i)	MSE(100,i)	EVAR(100,i)	LAG-i
-.0205	100	.0155	.0158	0
.00888	100	.0384	.0538	1
.0296	100	.0566	.0600	2
-.00849	100	.0527	.0601	3
-.0189	100	.0548	.0601	4
.00353	100	.0527	.0602	5
.0131	100	.0586	.0603	6
-.0133	100	.0710	.0603	7
-.00213	100	.0667	.0602	8
.00221	100	.0625	.0599	9
.00520	100	.0611	.0596	10
-.00469	100	.0631	.0597	11
-.0177	100	.0553	.0591	12
.0153	100	.0497	.0576	13
.00530	97	.0531	.0537	14
.00225	91	.0502	.0444	15
-.0259	72	.0435	.0334	16
.0147	54	.0344	.0194	17
.0370	29	.0144	.00444	18

Table 2.8 cont.

Lag Distribution (d)

T=200

General Statistics

$m^*(200) = 20.75$ $\hat{\sigma}^2(200) = .9959$ $K\text{-}S(200) = .2566$

Coefficient Statistics

BIAS(200,i)	NTIME(200,i)	MSE(200,i)	EVAR(200,i)	LAG-i
-.0138	100	.00676	.00620	0
.0181	100	.0251	.0216	1
.00589	100	.0302	.0241	2
-.00831	100	.0221	.0242	3
-.00143	100	.0234	.0242	4
.00269	100	.0238	.0241	5
-.00505	100	.0277	.0241	6
-.00406	100	.0315	.0241	7
.0113	100	.0251	.0242	8
-.00942	100	.0241	.0243	9
.00944	100	.0268	.0242	10
-.0169	100	.0247	.0242	11
.0105	100	.0212	.0242	12
.00704	100	.0270	.0241	13
-.00173	100	.0296	.0240	14
-.00906	100	.0271	.0234	15
-.0143	97	.0243	.0224	16
.0338	92	.0219	.0203	17
-.0243	83	.0214	.0170	18
.0209	67	.0193	.0137	19
-.00869	54	.0162	.0113	20
-.00124	46	.0167	.00916	21
-.00750	37	.00970	.00746	22
.0116	30	.00959	.00600	23
-.000980	23	.0101	.00494	24
-.0123	20	.00641	.00387	25
-.000397	15	.00744	.00271	26
.000952	11	.00327	.000710	27

Table 2.9 - Lag Distribution (d)

T= 50

Number of replications in each probability cell

	1	2	3	4	5	6	7	8	9	10	11	12	13
	Probabilities												
Lag	0.0-.0125	.0125-.025	.025-.05	.05-.1	.1-.3	.3-.5	.5	.5-.7	.7-.9	.9-.95	.95-.975	.975-.9875	.9875-1.0
0	0	0	0	0	0	0	0	0	0	1	0	2	97
1	0	0	0	0	0	0	0	4	13	10	11	10	52
2	0	0	0	0	0	2	0	10	22	10	16	12	28
3	0	0	0	1	1	9	0	18	30	19	5	3	14
4	0	0	0	0	3	10	0	16	32	14	11	3	11
5	1	0	0	1	6	12	0	21	34	10	8	2	5
6	0	0	0	0	8	13	0	21	30	14	3	5	7
7	1	0	0	3	14	13	0	21	31	4	5	2	6
8	1	0	0	5	12	14	0	18	29	10	3	2	6
9	1	2	0	4	12	26	1	20	20	6	3	1	4
10	0	2	0	4	14	12	2	26	20	8	3	0	9
11	0	3	4	7	21	18	12	8	7	2	2	3	13
12	0	0	0	0	0	0	32	3	0	7	13	11	34

T=100

0	0	0	0	0	0	0	0	0	0	0	0	0	100
1	0	0	0	0	0	0	0	0	0	3	2	2	93
2	0	0	0	0	0	1	0	0	6	5	7	13	68
3	0	0	0	0	0	1	0	5	16	8	16	8	46
4	0	0	0	0	0	5	0	10	21	13	13	16	22
5	0	0	0	0	0	10	0	10	22	18	11	12	17

Table 2.9 - Lag Distribution (d)

T=100

Number of replications in each probability cell

	1	2	3	4	5	6	7	8	9	10	11	12	13
	Probabilities												
Lag	0.0-.0125	.0125-.025	.025-.05	.05-.1	.1-.3	.3-.5	.5	.5-.7	.7-.9	.9-.95	.95-.975	.975-.9875	.9875-1.0
6	0	0	0	0	6	7	0	11	31	16	6	9	14
7	0	0	0	2	10	12	0	16	31	9	6	5	9
8	0	0	1	0	13	17	0	10	28	7	12	3	9
9	0	0	2	2	12	15	0	17	23	14	8	4	3
10	0	0	0	3	15	20	0	19	19	8	4	4	8
11	0	1	2	4	14	17	0	23	22	6	6	2	3
12	0	0	0	5	21	18	0	18	22	9	3	2	2
13	1	1	0	1	17	17	0	24	25	4	4	2	4
14	0	1	1	4	18	20	3	16	22	5	1	1	8
15	0	0	4	7	15	11	9	13	18	8	5	5	5
16	0	1	4	3	17	12	28	10	7	2	2	6	8
17	1	0	0	4	8	10	46	3	5	1	6	7	9
18	0	0	0	0	3	1	71	0	6	3	3	4	9

T=200

0	0	0	0	0	0	0	0	0	0	0	0	0	100
1	0	0	0	0	0	0	0	0	0	0	0	0	100
2	0	0	0	0	0	0	0	0	0	0	4	1	95
3	0	0	0	0	0	0	0	0	2	0	8	4	86
4	0	0	0	0	0	0	0	0	8	12	4	11	65
5	0	0	0	0	0	1	0	4	16	9	12	11	47
6	0	0	0	0	4	4	0	10	20	7	10	14	32

Table 2.9 - Lag Distribution (d)

T=200

Number of replications in each probability cell

	1	2	3	4	5	6	7	8	9	10	11	12	13
	Probabilities												
Lag	0.0-.0125	.0125-.025	.025-.05	.05-.1	.1-.3	.3-.5	.5	.5-.7	.7-.9	.9-.95	.95-.975	.975-.9875	.9875-1.0
7	0	0	0	0	5	6	0	13	27	11	6	8	24
8	0	0	0	1	5	8	0	12	25	13	15	8	12
9	0	0	2	0	9	9	0	17	33	8	10	5	7
10	0	2	1	0	9	9	0	19	28	12	8	7	5
11	1	0	1	1	13	11	0	26	28	8	4	3	4
12	0	0	0	2	13	11	0	25	28	7	6	4	4
13	0	0	2	4	17	12	0	23	17	10	7	4	4
14	1	0	3	3	18	16	0	19	18	11	3	4	4
15	2	1	2	1	21	18	0	18	22	6	2	2	5
16	0	2	3	6	17	15	3	23	18	5	3	1	4
17	0	0	2	1	15	13	8	19	22	4	4	5	7
18	0	2	5	4	14	17	17	12	12	8	2	1	6
19	1	0	1	2	9	11	33	12	12	5	5	3	6
20	2	2	0	3	13	6	46	3	13	4	3	1	4
21	1	4	1	1	5	9	54	5	7	3	3	0	7
22	0	1	1	3	9	5	63	5	6	3	2	1	1
23	1	0	0	1	5	1	70	5	7	4	1	3	2
24	2	1	1	2	1	3	77	1	8	0	1	2	1
25	2	0	0	1	7	2	80	4	0	2	1	0	1
26	0	0	2	1	4	0	85	2	2	0	1	0	3
27	3	0	0	0	0	1	89	2	1	0	0	0	4

For each replication let m* denote the lag length selected by the CAT criterion and let M denote the true lag length. The results of the Monte Carlo experiments indicate that when the CAT criterion is applied to the truncated lag distributions a and c, m* is greater than or equal to M except for 4 replications of lag distribution c, sample size 50, and m* is equal to M for approximately half the replications. The sample probabilities that m* = M+i, i=0, 1, ..., ($T^{.6}$-M) roughly correspond to the limiting probabilities that m* = M+i, i=0, 1, ..., 14 which were tabulated in section two, page 26. As one would expect, there is greater coincidence between the sample and limiting probabilities the larger is the sample size.

The coefficients in lag distribution b decline linearly from one to zero in steps of .10. Coefficients whose population value is close to zero are frequently excluded from the lag distribution chosen by the CAT criterion, although the frequency of replications where m* is less than M declines as sample size increases. A similar result holds for the infinite geometric lag distribution d. In this experiment, the average m* is an increasing function of the sample size. Although we have not analyzed the large sample properties of the CAT criterion when the population distributed lag is infinite, the results of experiment d indicate that the bias of the estimated coefficients is small despite the specification error.

For experiments a, b, and c the null hypothesis that the sum of the estimated coefficients is equal to the sum of the population coefficients is easily accepted at a 5% significance level. This hypothesis is rejected for experiment d (all sample sizes) at any reasonable significance level. The latter result is not surprising as the fitted lag distributions are too short. For sample sizes of 50, 100, and 200, the average m* is equal to 11.5, 16.4, and 20.8, and the difference between the sum of experiment d population distributed lag coefficients (5.0) and the sum of the first 11, 16, and 21 population coefficients is .43, .14, and .05. This analysis suggests that the majority of the ratios (the sum of the estimated coefficients minus the actual sum divided by the standard error of the sum) used in the calculation of the Komolgorov-Smirnov statistics are positive. So it is unlikely that these sample ratios for experiment d are realizations from a N(0,1) population and hence the null hypothesis is rejected.

There is a discrepancy between the mean square error (MSE) of the estimated coefficients for all experiments and sample sizes and the average estimated variance of the coefficients, EVAR. This discrepancy has two components. First, one can only expect the approximate equality of MSE and EVAR when the coefficient bias is zero. The average bias of the estimated coefficients over all experiments is small, so the coefficient bias explains only a part of the MSE-EVAR disparity. Second, the average estimate of σ^2, the variance of the disturbance term, is biased downwards for experiments a, b, and c and upwards for experiment d, except for experiment d, T=200, where σ^2 is essentially unbiased. In all cases the bias tends to zero as sample size increases. The direction of the σ^2 bias is as expected since the average m* exceeds M for the finite distributed lag models a, b, and c, and m* is always too short for the infinite distributed lag model d. For those coefficients which are always included in the lag distribution selected by the CAT criterion, experiments a-c, EVAR is usually less than MSE. The reverse is true for experiment d except for the case T=200 when EVAR and MSE are roughly coincidental. These results reflect the direction of the σ^2 bias. Last, for those coefficients which are frequently omitted from the lag distribution selected by the CAT criterion, all experiments, there is greater disparity between EVAR and MSE, the smaller is NTIME.

We now turn our attention to the analysis of tables 2.3, 2.5, 2.7, and 2.9. As was stated earlier, the null hypothesis that the ratio of an estimated coefficient to its estimated standard error is distributed as a N(0,1) is incorrect for the first M+1 coefficients of each lag distribution and correct for all others. When the null hypothesis is true, one would expect to find a distribution of the F(i,T,k) across the columns of tables 2.3, 2.5, 2.7, and 2.9 in proportion to the probabilities at the head of each column. Since a different order lag distribution is selected each replication, and since those coefficients not included in the fitted model are assigned a cumulative probability of .5, we cannot expect the F(i,T,k) to be distributed across the columns of these tables in the manner described above -- except for those coefficients just beyond the end of the true lag distribution which are frequently included in the fitted model, but whose population value is zero. Despite the drawbacks of this analysis, it is still possible to make general statements concerning the type I and type II error probabilities associated with the maintained hypothesis $\beta(i) = 0$, for all i, when the CAT criterion selects the order of the estimated distributed lag model.

A review of tables 2.3, 2.5, 2.7, and 2.9 indicates that for the first M+1 coefficients, the probability of a type II error (accepting the hypothesis $\beta(i) = 0$, $i=0, \ldots, M-1$ when it is false) decreases with sample size and increases as one moves closer to the end of the true lag distribution. To see this note that the number of replications in the last column of tables 2.5, 2.7, and 2.9 for the first M+1 coefficients increases (to a maximum of 100) as sample size increases, but the number of replications in the last column decreases as one moves closer to the end of the true lag distribution. This generality does not apply to the first experiment, lag distribution a (M=0), since in this case there are 100 replications in the first row and last column of the table 2.3 for all sample sizes. For those estimated coefficients whose population value is zero, the probability of a type I error (rejection of the null hypothesis $\beta(i) = 0$, $i=M+1, \ldots$ when it is true) tends to decline as sample size increases, and as one moves further away from the end of the true lag distribution. To see this note that the number of replications in column 7 of tables 2.3, 2.5, 2.7 and 2.9 for the M+1 through $(T^{.6})$-th coefficients increases with sample size, and is larger (with a maximum of 100) for those coefficients corresponding to the longest lags. In summation, it is only for those estimated coefficients in a band around the true lag length M that the coefficient hypothesis tests tend to be biased, and it is quite likely that EVAR-MSE discrepancy is a principle source of this bias.

The results of this section constitute strong evidence for the use of the CAT criterion to estimate the order and the coefficients of the distributed lag model 1.3. Using moderate size samples we have found corroborative evidence for the limiting probabilities that $m^* = M+i$, $i=0, 1, \ldots, M$ derived in section two. Despite the similarity of the CAT criterion to regression strategies (18, pp. 603-606), coefficient t-statistics (except for those coefficients in a band round the true lag length M) are not biased by the use of the CAT selection procedure. But until we derive the large sample properties of Parzen's criterion for the case of infinite lag distributions, the applicability of the CAT criterion is limited to circumstances where the researcher has a priori knowledge that the lag distribution is finite.

Footnotes

[1]See Kmenta (10), pages 282-294 or Johnston (9), pages 259-265 for a discussion of these procedures.

[2]Sims (17) provides an excellent survey of the various estimators of the distributed lag model. He shows (pp. 305-308, 326-329) that there exists a sequence of m's converging to inifinity with T, $m/T \to 0$, so that ordinary least squares, feasible generalized least squares, and Hannan inefficient (7) estimators of the DLM all have the same asymptotic distribution.

[3]In spectral theory, see Hannan (8, pp. 273-288) for a discussion of the relationship between expanding parameterizations and sample size. Amemiya (3) does not provide guidelines for choosing the order of the residual autoregression as a function of the sample size, but the ideas presented in Hannan still applicable.

[4]Theoretically, the CAT criterion selects the order of an approximating autoregressive process which minimizes the one step ahead mean square prediction error. In (12) a set of AR models are estimated using monthly economic data (1960-1974) to see if the CAT criterion selects the appropriate order of an AR process. These experiments are inconclusive, but they suggest that the CAT decision rule does not always select AR models that minimize mean square prediction error.

[5]See Hannan (8), pages 204-220, especially theorem 6.

[6]The matrix result $A \geq B > 0 \Rightarrow B^{-1} \geq A^{-1}$ is well known; see Goldberger (6 , p. 38).

[7]Anderson shows that $\frac{1}{\sqrt{T}} \; X'_m \underline{\varepsilon}$ has a limiting normal distribution with zero mean vector and covariance matrix $\sigma^2 Q(m)$ given assumptions that are satisfied by our covariance stationary x process. His theorem 2.6.1 (4, pp. 23-24) implies that

$\operatorname*{plim}_{T\to\infty} \frac{1}{T} X'_m \underline{\varepsilon} = \underline{0}$ for m fixed.

[8]See Theil (18, p. 380). When $m < M$, $\operatorname*{plim}_{T\to\infty} \sigma^2_m$ is strictly greater than σ^2 because $\underline{\beta}_M \neq \underline{0}$ and

$$\underset{T\to\infty}{\text{plim}} \left(\frac{X'_M P_{m^*} X_M}{T}\right) = \left(\begin{array}{c|c} 0 & 0 \\ \hline 0 & Q^*(M-m)-Q^*(m,M-m)'Q^*(m)^{-1}Q^*(m,M-m) \end{array}\right),$$

and the lower right hand block of the limit matrix has full rank M-m.

[9] Lemma 1: Let s(t) be a sequence of random variables, $s(t) \geq 0$ for all t, $E(s(t)) < \infty$ for all t. If $\lim_{T\to\infty} E(s(t)) = 0$, then $\underset{T\to\infty}{\text{plim}}\, s(t) = 0$.

Proof: Let $\delta > 0$, and let f(s(t)) denote the probability density function (pdf) of s(t). Then

$$E(s(t)) = \int_0^{\delta} s(t)f(s(t))ds(t) + \int_{\delta}^{\infty} s(t)f(s(t))ds(t),$$

$$E(s(t)) \geq \delta \cdot \int_{\delta}^{\infty} f(s(t))ds(t),$$

$$\left(\frac{1}{\delta}\right)E(s(t)) \geq \int_{\delta}^{\infty} f(s(t))ds(t) = \text{Prob}(s(t) \geq \delta).$$

Consequently

$$\left(\frac{1}{\delta}\right)\lim_{t\to\infty} E(s(t)) \geq \lim_{t\to\infty} \text{Prob}(s(t) \geq \delta) = \lim_{t\to\infty} \text{Prob}(|s(t)-0| \geq \delta),$$

thus

$$0 \geq \lim_{t\to\infty} \text{Prob}(|s(t) - 0| \geq \delta), \text{ that is: } \underset{T\to\infty}{\text{plim}}\, s(t) = 0.$$

[10] If a function $m^o(T)$ had been chosen so that $m^o(T)\to\infty$ as $T\to\infty$ while $\lim_{T\to\infty} \frac{m^o(T)}{T} > 0$-- for example $m^o(T) = b\cdot T$, $0 < b < 1$-- then the arguments presented on pages 16-17 cannot be used to show $\underset{T\to\infty}{\text{plim}}\, CAT(m^o(T)) = -\sigma^{-2}$. This is so because

$$\lim_{T\to\infty} E \frac{\varepsilon' N_m^o(T)\,\varepsilon}{T} = \sigma^2 b,$$

and the conditions for lemma 1 footnote 9 are no longer satisfied.

[11] I am indebted to John Geweke, my thesis advisor, for his help in deriving the results on pages 19-25.

[12] For n events $A_1, \ldots, A_n$ one uses the definition of conditional probability and an inductive argument to show

$$P(\bigcap_{i=1}^{n} A_i) = P(A_1)P(A_2|A_1)P(A_3|A_2 \cap A_1)\ldots P(A_n| \bigcap_{i=1}^{n-1} A_i).$$

There are n! such formulae.

[13] The mechanics on pages 20-22 can also be used to analyze the large sample properties of the residual variance model fitting criterion, Theil (18, pp. 543-5): the expression $T \cdot (\hat{\sigma}^2_{M+k} - \hat{\sigma}^2_M)$ converges in distribution to $\sigma^2(k - \chi^2(k))$. The

$$\lim_{T\to\infty} \mathrm{Prob}(m^* = M) = \prod_{i=1}^{\infty} R_i,\ R_i = \mathrm{Prob}(u_{i-1} + v < i|\ u_{i-1} < (i-1)),$$

$i=1, \ldots,$ where u_{i-1} and v are defined as they were in the text. In this case, the limiting probability that $m^*=M$ is not greater than zero because the R_i converge too slowly to one. To see this note that the expectation of $\sigma^2(j-\chi^2(j))$ is zero and the variance of $\sigma^2(j-\chi^2(j))$ is $2\sigma^4 j$. The convergence of R_j to one as $T\to\infty$ is too slow for the product of the R_j to be bounded away from zero. The probability that $\sigma^2(j-\chi^2(j))$ is greater than zero approaches .5 as j gets large. For the CAT criterion, $E(\frac{1}{\sigma^2}(2j-\chi^2(j)) = j/\sigma^2$, an increasing function of j. The probability that $\frac{1}{\sigma^2}(2j-\chi^2(j))$ is greater that zero goes to one as $j\to\infty$. Thus the convergence of the ρ_i to one is faster for the CAT criterion, and $\lim_{T\to\infty} \mathrm{Prob}(m^* = M) > 0$.

[14] The statement in the text must be interpreted with care. If x(t) is a white noise process, there is no gain in asymptotic efficiency over HI for the first M+1 coefficients since

$\operatorname{plim}_{T\to\infty} (\frac{1}{T}(X_m'X_m))^{-1}$, any fixed $m \geq M$, is a diagonal matrix. If x(t) follows an ARMA(p,q) process, then there is a gain in asymptotic efficiency over HI for some of the first M+1 coefficients, and it is the form of $\operatorname{plim}_{T\to\infty} (\frac{1}{T}(X_m' X_m))^{-1}$, any fixed $m \geq M$, which determines the coefficients with smaller asymptotic variance.

[15] We must first establish that $\operatorname{plim}_{T\to\infty} \left(\frac{X_M' P_{m(T)-M} X_M}{T} \right)$ exists, and then that it is positive definite. Suppose we look at a sequence of non-zero quadratic forms in the matrices 1.27 of the text, page 14. The limit of this sequence exists since the quadratic forms are monotone decreasing and bounded below.

Proposition F1: As $T\to\infty$, the limit of the sequence of non-zero quadratic forms in the matrices of 1.27 is greater than zero, i.e., $\operatorname{plim}_{T\to\infty} \frac{X_M' P_{m(T)-M} X_M}{T}$ is positive definite.

Proof: Suppose the x process has moving average representation,

$$x(t) = \sum_{s=0}^{\infty} b(s)\varepsilon(t-s),\ b(0) \equiv 1, \text{ and } \sum_{s=0}^{\infty} b(s)^2 < \infty.$$

The matrix $C = \operatorname{plim}_{T\to\infty} (X_M' P_{m(T)-M} X_M/T)$ can be interpreted as the matrix of 1 to (M+1) step ahead prediction error variances and covariances from a projection of x(t) on its infinite past history. Let $\Omega(t)$ be the set of observed innovations at time t, $\Omega(t) = \{\ldots\ \varepsilon(t-1),\ \varepsilon(t)\}$. The i^{th} diagonal element of C, C(i,i), is equal to

$$E(x(t+M+2-i)^2|\Omega(t)) = E\left[(\sum_{s=0}^{\infty} b(s)\varepsilon(t+M+2-i-s))^2|\Omega(t)\right]$$

$$= \sum_{s=0}^{\infty} b(s)^2\, E\left[(\varepsilon(t+M+2-i-s))^2|\Omega(t)\right] = \sigma^2 \sum_{s=0}^{M+2-i} b(s)^2,$$

for i=1, ..., M+1. C(i,i) is the (M+2-i) step ahead prediction error variance of x(t). The covariances C(i,j) = min {C(i,i), C(j,j)}. Let $\underline{s}$ be a (M+1) x 1 vector, $\underline{s} \neq \underline{0}$. Suppose $\underline{s}'C\underline{s} = 0$. This implies that there exist some j, $0 < j \leq M+1$ such that x(t+j) is perfectly predictable (with probability one) for all t. This is impossible since x(t) is nondeterministic. Therefore, C is positive definite. The conclusion in the text is appropriate since the inverse of C is a continuous function of the elements of C.

[16] The standard normal variates are termed pseudo-random since they are generated by the method of Box and Muller (5) on a Univac 1110 digital computer at the University of Wisconsin, Madison.

[17] Most quarterly economic time series are well represented by stochastic second order difference equations, see Sargent (16, chapter XI).

[18] An infinite geometric lag distribution is not appropriate for model 1.3; we include this parameterization in our set of experiments to gain insight into the behavior of the CAT criterion when applied to an infinite lag distribution.

[19] Despite the fact that the disturbance term of 1.3 is distributed as a N(0,1), we have not shown that the estimated coefficient vector selected by the CAT criterion is normally distributed. In empirical work it is convenient to assume that the order of the fitted model is fixed, and to proceed with conventional hypothesis tests as if they were appropriate. The results of the Monte Carlo experiments reported in the text indicate that this simplification does not result in test statistics which are grossly distorted.

[20] We chose to analyze the null hypothesis that the sample ratios of the estimated coefficients to their estimated standard errors are distributed as a N(0,1) although it is incorrect for the first M+1 coefficients, since the magnitude of the t ratio is more often than not the decision criterion used by empirical researchers in determining what variables to keep in their models.

Bibliography

1) Aitken, A.C. (1935) "On Least Squares and Linear Combinations of Observations." Proceedings of the Royal Society of Edinburgh, 55, pp. 42-48.

2) Akaike, H. (1974), "A New Look at the Statistical Model Identification," IEEE Trans. Auto. Control, Vol. AC-19, pp. 716-723.

3) Amemiya, T. (1973), "Generalized Least Squares with an Estimated Autocovariance Matrix," Econometrica, 41, No. 4, 723-732.

4) Anderson, T.W. (1971), The Statistical Analysis of Time Series, New York: John Wiley and Sons.

5) Box, G.E.P., and Muller, M.E., "A Note on the Generation of Random Normal Deviates," Ann. Math. Statistics, 28 (1958), pp. 610-611.

6) Goldberger, A.S., (1964), Econometric Theory, New York: John Wiley & Sons.

7) Hannan, E.J. (1963), "Regression for Time Series," in Proceedings of a Symposium on Time Series Analysis, M. Rosenblatt (ed.), New York, John Wiley and Sons.

8) Hannan, E.J. (1970), Multiple Time Series, New York: John Wiley and Sons.

9) Johnston, J. (1963), Econometric Methods, New York: McGraw-Hill Book Company Inc.

10) Kmenta, J. (1971), Elements of Econometrics, New York: Macmillan Publishing Co., Inc.

11) Lindgren, B.W., Statistical Theory, London: Macmillan, 1968.

12) Meese, R., "Distributed Lag Order Determination With an Application to the Multiperiod Theory of the Firm" unpublished Ph.D. thesis, University of Wisconsin, Madison, 1978.

13) Parzen, E. (1974), "Some Recent Advances in Time Series Analysis," IEEE Trans. Auto. Control, Vol. AC-19, pp. 723-730.

14) __________(1975), "Multiple Time Series: Determining the Order of Approximating Autoregressive Schemes," Technical Report 23, State University of New York (SUNY) at Buffalo.

15) __________ (1976) "An Approach to Time Series Modeling and Forecasting Illustrated by Hourly Electricity Demands," Technical Report No. 37, SUNY at Buffalo.

16) Sargent, T., Notes on Macroeconomic Theory, University of Minnesota, 1977.

17) Sims, C. (1974), "Distributed Lags," in M. Intriligator and D. Kendrick, (eds.), Frontiers of Quantitative Economics, Volume II, Amsterdam, North Holland.

18) Theil, H. (1971), Principles of Econometrics, New York: John Wiley and Sons.