ECONOMIC RESEARCH REPORTS

THE IDENTIFIABILITY AND ESTIMABILITY OF NON-INVERTIBLE MA(Q) MODELS

by

James B. Ramsey

and

Alvaro Montenegro

R.R. #88-08

March 1988

C. V. STARR CENTER FOR APPLIED ECONOMICS

NEW YORK UNIVERSITY
FACULTY OF ARTS AND SCIENCE
DEPARTMENT OF ECONOMICS
WASHINGTON SQUARE
NEW YORK, N.Y. 10003

THE IDENTIFIABILITY AND ESTIMABILITY OF NON-INVERTIBLE MA(Q) MODELS

James B. Ramsey
Alvaro Montenegro
Economics Department
New York University
March 20, 1988

ABSTRACT

The second order properties of a given MA process are compatible with several different model specifications, only one of which is invertible. Consequently, traditional estimation procedures, such as OLS or Box and Jenkins ARIMA modelling, which are based on second order properties, are incapable of distinguishing among these alternative model specifications; this ambiguity is usually resolved by restricting the process to be of the invertible type. This paper presents an estimation procedure, based on higher order moments, which is capable of distinguishing between these alternative specifications, without recourse to the invertibility assumption. The true sequence of innovations that drive the MA process can be estimated once the correct model is determined. Also discussed, is the finding that the application of OLS to a non-invertible MA process generates an ARCH structure in the residuals.

March 20,1988

THE IDENTIFIABILITY AND ESTIMABILITY OF NON-INVERTIBLE MA(Q) MODELS

James B. Ramsey
Alvaro Montenegro
Economics Department
New York University

It is well known that for any MA(q) stochastic process, there corresponds 2^q alternative models that are not distinguishable from each other using the usual autocorrelation based techniques (OLS or Box and Jenkins ARIMA procedures, for example). Of these 2^q models, only one is invertible. This is the only model estimated as it is the one model that can be re-expressed as a weighted lag of the past observed members of the time series. In this paper, we demonstrate the conditions under which we can identify the other 2^q-1 models, estimate their parameters, and estimate the unobserved underlying innovations. In the process, we demonstrate that the estimation of a noninvertible model by traditional means generates an ARCH process in the residuals.

Some of the issues that we address in this paper within the context of the time domain have been discussed by Lii and Rosenblatt (1982) within the context of the frequency domain; further references are cited in the bibliography. The advantage of the time domain approach is that it allows us to concentrate on the structure of the models and to relate the results to the precise formulation of those models, a task that is particularly appealing to economic analysts.

The outline of this paper is simple. The first section recalls the proof that there are 2^q models that have the same auto-correlation function so that they are not distinguishable by traditional means. All of these models provide the same linear forecast. The second section of the paper demonstrates that non-normality of the innovations can be used in conjunction with higher order moments to identify and to estimate the parameters of each of the other models. In the process, the "pathology" of the normal distribution is shown; there exist 2^q distinct models as defined by their parameter values, but as far as we can discover all of these models are statistically indistinguishable from each other, even though the time path of each model is different. The next section utilizes the results of the previous section to provide a test for identifiability, to demonstrate the estimation of the original innovations, and to show that the standard parameter estimation procedure, when applied to a non-invertible model leads to the generation of residuals with an ARCH structure. In this context, a beginning is made on the use of our ability to estimate the innovations in trying to learn more about the total statistical properties of the stochastic process. There is considerable unused information in the original innovations that has not been exploited. The fourth section briefly indicates that indeed there are actual economic time series that are sufficiently non-normal to enable us to use our new tools. Illustrations of the theoretical concepts discussed are provided in each section. The paper is concluded with some summary comments and some ideas for further research.

(1) THE EQUIVALENCE OF MA(q) MODELS

We consider the following linear model;

$$X_t = \Sigma_0^q \alpha_s \epsilon_{t-s}; \tag{1.1}$$

$$= \alpha_0 \epsilon_t + \alpha_1 \epsilon_{t-1} + \alpha_2 \epsilon_{t-2} + \ldots . \alpha_q \epsilon_{t-q}.$$

The distribution of the $\{\epsilon_t\}$ is assumed to be a white noise process in that terms in the sequence $\{\epsilon_t\}$ are mutually uncorrelated and have a zero mean and a constant variance of σ_ϵ^2. By solving the roots of the auxilary equation;

$$\alpha_0 + \alpha_1 B + \alpha_2 B^2 + \ldots + \alpha_q B^q; \tag{1.2}$$

we obtain the alternative form of expressing equation (1.1):

$$X_t = \Pi_1^q (1-\lambda_i B) \epsilon_t; \tag{1.3}$$

where "B" is the backward shift operator. The λ_i, $i = 1,2,\ldots q$, are the roots of the auxiliary equation (1.2). It is well known that invertibility requires that the maximum modulus of the set of roots be less than one in absolute value. The reader will recall that invertibility in the linear model means that equation (1.1) can be "inverted" to yield:

$$\epsilon_t = \Sigma_0^\infty \beta_i X_{t-i}; \tag{1.4}$$

where the infinite series is convergent for almost all realizations of the stochastic process $\{X_t\}$.

Let $\{\lambda_i\}$ denote the set of roots corresponding to some one choice of model coefficients as defined by the choice of a coefficient set $\{\alpha_i\}$, i = 1,2, ..q, where, without loss of generality, we use the normalization rule that $\alpha_0=1$. Let $\{\lambda_i^{'}\}$ denote an alternative set of roots that differ from the first set only in that for some values of k, $k\in\{1,2,...q\}$, λ_k is replaced by λ_k^{-1}. Corresponding to this choice of set of roots, there is a set of model coefficients $\{\alpha_i^{'}\}$ that differs from the original set of coefficients $\{\alpha_i\}$.

The two models defined by $\{\lambda_i\}$ and $\{\lambda_i^{'}\}$, or by $\{\alpha_i\}$ and $\{\alpha_i^{'}\}$, are autocorrelation equivalent in that both models have the same autocorrelation function. This well known result can easily be demonstrated by using the concept of the autocovariance generating function; see for example, Brockwell and Davis (1987). We define the autocovariance generating function by:

$$G(z) = \Sigma_{k=-\infty}^{\infty} \gamma(k)z^k; \tag{1.5}$$

with convergence in some annulus $r^{-1}<|z|<r$, $r>1$, where $\gamma(k)$ is the autocorrelation function. If X_t is an MA(q) process, then G(z) can be rewritten as;

$$G(z) = \sigma_\epsilon^2\alpha(z)\alpha(z^{-1}); \tag{1.6}$$

where $\alpha(z)$ is defined by:

$$\alpha(z) = 1 + \alpha_1 z + \alpha_2 z^2 + \alpha_3 z^3 \; \ldots\ldots \alpha_q z^q;$$

and equation (1.6) follows from substituting the definition of $\gamma(k)$ into equation (1.5) and the fact that for a real process the autocorrelation function is an even function of the index; that is, $\gamma_j = \gamma_{-j}$.

We conclude that if z_0 is a root of equation (1.6), then so is z_0^{-1}. There are 2^q choices of q roots from the 2q roots, $\{\lambda_i\}$ and $\{\lambda_i^{-1}\}$, since for each distinct root there are two choices of associated power, ±1. Consequently, any choice of q values from the set $\{\lambda_i, \lambda_i^{-1}\}$ will yield upto a scale factor the same autocovariance generating function and therefore exactly the same set of auto-correlations. Indeed, one might well characterize the auto-correlation function as being "time blind".

This result also shows that any property that relies exclusively on the auto-correlation function will be the same for all members of the auto-correlation equivalent group of models. For example, the linear one-step ahead OLS forecast $\hat{X}_T$, is given by:

$$\hat{X}_T = -\alpha_1 X_{T-1} - \alpha_2 X_{T-2} - \alpha_3 X_{T-3} \ldots\ldots \qquad (1.7)$$

the statistical properties of which depend only on the auto-correlation function of the observed sequence $\{X_t\}$. Consequently, all these auto-correlation equivalent models yield precisely the same forecast.

As an example we present the following four MA(2) models, with $\{\lambda_i\}=\{2,3\}$, that exhibit the same autocorrelation function: -.565 for lag 1 and .097 for lag 2. The last model, (1.8d), corresponds to the roots $\{\lambda_i\}=\{2,3\}$. The invertible model is (1.8c).

EQUATION	ROOTS	
$x_{1t} = \epsilon_t - 2.33\epsilon_{t-1} + .666\epsilon_{t-2}$	(2,1/3)	(1.8a)
$x_{2t} = \epsilon_t - 3.5\epsilon_{t-1} + 1.5\epsilon_{t-2}$	(1/2,3)	(1.8b)
$x_{3t} = \epsilon_t - .833\epsilon_{t-1} + .1666\epsilon_{t-2}$	(1/2,1/3)	(1.8c)
$x_{4t} = \epsilon_t - 5\epsilon_{t-1} + 6\epsilon_{t-2}$	(2,3)	(1.8d)

(2) THE IDENTIFICATION OF MODELS WITH NON-NORMAL INNOVATIONS

In this section we demonstrate that for models based on non-normally distributed innovations, that is, for models where $\{\epsilon_t\}$ has a distribution other than normal, the 2^q distinct models can be identified and their parameters estimated consistently. For the corresponding approach using the polyspectra, see Lii and Rosenblatt (1982), or Matsuoka and Ulrych (1984) with associated references.

The assumed model is that presented in equation (1.1), except that normality of the probability distribution function is specifically denied. However, we do assume stationarity up to the fourth order, i.e., all moments up to the fourth are constant. In the final part of this section we will discuss the difficulties with the normal distribution.

Equations (1.1) and (1.3) summarize the two ways of writing out the Ma(q) process, one in terms of the model coefficients and one in terms of the roots of the equation. The roots and the coefficients are related by the following expressions for an MA(q) process.

$$\alpha_1 = (-1)\Sigma_i \lambda_i$$

$$\alpha_2 = (-1)^2 \Sigma_{j>i}^{q} \Sigma_{i=1}^{q-1} \lambda_i \lambda_j$$

$$\alpha 3 = (-1)^3 \Sigma_{k>j}^{q} \Sigma_{j>i}^{q-1} \Sigma_{i=1}^{q-2} \lambda_i \lambda_j \lambda_k$$

$$\vdots$$

$$\alpha_q = (-1)^q \lambda_1 \lambda_2 \lambda_3 \ldots . \lambda_q \qquad (2.1)$$

The statement of the main theorem is simple. The 2^q distinct models generated by alternative choices of $\lambda_i^{\pm 1}$, where $\{\lambda_i\}$ are the q roots for some choice of q model coefficients, $\{\alpha_i\}$, can be identified, if, and only if, the stationary distribution of the sequence $\{\epsilon_t\}$ is not normally distributed, provided only that the joint distribution of $\{\epsilon_t\}$ has finite moments up to the fourth order.

The proof is by induction. The procedure is first to show that if an MA model of order q-1 is identified, then a model of order q is identified. The next step is to show that an MA(1) model is identified when the ϵ_t are not

Gaussian. The last step is to show that the models are not identified by the procedures used in the first part of the proof when the $\{\epsilon_t\}$ are normally distributed.

We begin by defining the bicovariances for any MA process, where the innovation sequence $\{\epsilon_t\}$ has zero mean see, for example, Rao and Gabr (1984, pg.118).

$$\begin{aligned} Bc(t_1,t_2) &= E\{X_t X_{t+t_1} X_{t+t_2}\} \\ &= \Sigma_u \Sigma_v \Sigma_w \alpha_u \alpha_v \alpha_w E\{\epsilon_{t-u}\epsilon_{t+t_1-v}\epsilon_{t+t_2-w}\} \\ &= \mu_3(\epsilon)[\Sigma_u \alpha_u \alpha_{u+t_1}\alpha_{u+t_2}] \end{aligned} \tag{2.2}$$

Consider the subset of bicovariances of lag length q, max $|t_i|=q$, i=1,2, in order to avoid sums of the products of the model coefficients. The terms are:

$$\begin{aligned} Bc(\ 0,-q) &= (\mu_3\alpha_0\alpha_q)\alpha_q \\ Bc(-1,-q) &= (\mu_3\alpha_0\alpha_q)\alpha_{q-1} \\ Bc(-2,-q) &= (\mu_3\alpha_0\alpha_q)\alpha_{q-2} \\ Bc(-3,-q) &= (\mu_3\alpha_0\alpha_q)\alpha_{q-3} \\ &\ldots\ldots\ldots\ldots\ldots\ldots \\ Bc(-q,-q) &= (\mu_3\alpha_0\alpha_q)\alpha_0; \end{aligned} \tag{2.3}$$

or: $Bc(-j,-q) = K\alpha_{q-j}$, where $K = \mu_3\alpha_0\alpha_q$.

The bicovariances are all proportional to the individual model coefficients. This characteristic is the clue to model identification using the bicovariance.

Let $\{\lambda_i\}$, i=1,...,q-1, designate the set of roots for a (q-1) order model with model coefficients $\{\beta_i\}$, i=1,...,q-1, so that the q'th order model can be written in terms of the (q-1) roots $\{\lambda_i\}$ and the additional q'th root for the q'th order model, designated as λ_q. We can express the q'th order model's coefficients, $\{\alpha_i\}$, in terms of the β_i and the q roots, λ_i, by:

$$
\begin{aligned}
\alpha_1 &= \beta_1 - \lambda_q \\
\alpha_2 &= \beta_2 + \Sigma \lambda_i \lambda_q \\
\alpha_3 &= \beta_3 - \Sigma_i \Sigma_j \lambda_i \lambda_j \lambda_q \\
&\ldots\ldots\ldots\ldots\ldots\ldots \\
\alpha_{q-1} &= \beta_{q-1} + (-1)^{q-1} (\Sigma \ldots \Sigma \lambda_i \lambda_j \ldots \lambda_q) \\
\alpha_q &= \beta_{q-1}(-\lambda_q);
\end{aligned}
\tag{2.4}
$$

where the multiple summations on the λ_i at the k'th step run from 1 to q-k for index i, from i to q-k+1 for index j, from j to q-k+2 for the next index, and so on; the last summation is from the previous index to q. The alternative set of model coefficients, designated $\{\alpha'_i\}$, corresponding to the choice of roots $\{\lambda_i$, i=1,2,..q-1, $\lambda_q^{-1}\}$, is given by:

$$
\begin{aligned}
\alpha'_1 &= \beta_1 - \lambda_q^{-1} \\
\alpha'_2 &= \beta_2 + \Sigma_i \lambda_i \lambda_q^{-1} \\
\alpha'_3 &= \beta_3 - \Sigma_i \Sigma_j \lambda_i \lambda_j \lambda_q^{-1}
\end{aligned}
\tag{2.5}
$$

. . . .

$\alpha'_{q-1} = \beta_{q-1} + (-1)^{q-1}(\Sigma\ldots\Sigma\lambda_i\lambda_j\ldots\lambda_q^{-1})$

$\alpha'_q = \beta_{q-1}(-\lambda_{q1})$

Let us assume without loss of generality that the model set $\{\alpha_i\}$ are the model coefficients for the "true" or "correct" model and that the set $\{\alpha'_i\}$ are the coefficients for any other model. The task is to demonstrate that ratios of the α_i do not equal corresponding ratios of the α'_i. Ratios are examined because ratios of the bicovariances in (2.3) can provide estimates of the ratios of the model coefficients. If the ratios of the $\{\alpha'_i\}$ are not equal to the ratios of the $\{\alpha_i\}$, the coefficients of the assumed correct model, then, in principle at least, the bicovariances can be used to distinguish between the alternative models. The different models are identifiable, if $\alpha_i/\alpha_q \neq \alpha'_i/\alpha'_q$. This is easily verified. Consider, for example, the ratios α_1/α_q and α_2/α_q. The ratios are:

$$\alpha_1/\alpha_q = (\beta_1 - \lambda_q)/\ \beta_{q-1}(-\lambda_q); \qquad (2.6)$$

$$\alpha'_1/\alpha'_q = (\beta_1 - \lambda_q^{-1})/\ \beta_{q-1}(-\lambda_q^{-1}); \qquad (2.7)$$

by multiplying numerator and denominator of equation (2.7) by λ_q^2, we obtain:

$$\alpha'_1/\alpha'_q = (\lambda_q^2\beta_1 - \lambda_q)/\beta_{q-1}(-\lambda_q);$$

which demonstrates that, while the denominator is the same as that in equation (2.6), the numerator cannot be re-expressed to obtain the numerator of equation (2.6). Similarly with the ratio α_2/α_q, we obtain:

$$\alpha_2/\alpha_q = (\beta_2 + \Sigma\lambda_i\lambda_q)/ (\beta_{q-1}(-\lambda_q));$$
$$\alpha_2'/\alpha_q' = (\beta_2 + \Sigma_i\lambda_i\lambda_q^{-1})/(\beta_{q-1}(-\lambda_q^{-1}));$$

and again by multiplying both numerator and denominator by the appropriate power of λ_q, we can demonstrate that the models are not equivalent with respect to these higher moments. Similar arguments can be used to show that the result that we have just demonstrated for α_1 and α_2 is true for any pair (α_i,α_j). However, the proof so far depends on $\mu_3(\epsilon)$ being non-zero. To show the identifiability of the models when the third moment is zero, we will have to go to the tricovariances, which correspond to the fourth order moments.

It is convenient to introduce the cumulants as an alternative to moments. The cumulants are of great theoretical use, but are seldom used in estimation directly. The r'th cumulant, κ_r, is obtained by taking the k'th derivative of the log characteristic generating function, where the k'th moment is obtained by taking the k'th derivative of the characteristic generating function. The cumulants and moments are related by:

$$\begin{aligned}\mu_2 &= \kappa_2\\ \mu_3 &= \kappa_3\\ \mu_4 &= \kappa_4 + 3\kappa_2^2\end{aligned} \tag{2.8}$$

$$\begin{aligned}\kappa_2 &= \mu_2\\ \kappa_3 &= \mu_3\\ \kappa_4 &= \mu_4 - 3\mu_2^2\end{aligned} \tag{2.9}$$

For the normal distribution the cumulants beyond the second are all zero. Consequently, if the distribution of the innovations is normal, then the alternative models are completely unidentifiable by any means that rely on moments or cumulants.

When the third moment is zero we can no longer rely on the bicovariances in equation (2.3) for identification and must resort to the next higher order moment, the tricovariance in this case. Following Rao and Gabr (1984), the tricovariance, $Tr(t_1,t_2,t_3)$, can be expressed in terms of the cumulants as;

$$Tr(t_1,t_2,t_3) = c(t_1)c(t_3-t_2) + c(t_2)c(t_3-t_1) + c(t_3)c(t_2-t_1) + c(t_1,t_2,t_3) \qquad (2.10)$$

where $Tr(t_1,t_2,t_3) = E[X_t X_{t+t_1} X_{t+t_2} X_{t+t_3}]$.

$c(t)$ denotes the second order cumulant which is equal to the second order moment, the autocovariance, and $c(t_1,t_2,t_3)$ denotes the fourth order cumulant whose expression is

$$c(t_1,t_2,t_3) = (\mu_4 - 3\sigma_\epsilon^4)\ \Sigma\ \alpha_t \alpha_{t+t_1} \alpha_{t+t_2} \alpha_{t+t_3}. \qquad (2.11)$$

Equation (2.11) indicates immediately that this expression cannot be used to identify models in the Gaussian case because with the normal distribution

$\mu_4 = 3\sigma_\epsilon^4$ so that $c(t_1, t_2, t_3)$ is zero. We can re-express equation (2.10) in terms of the autocovariance, $R(t)$, and solve for the fourth order cumulant:

$$c(t_1,t_2,t_3) = Tr(t_1,t_2,t_3) - R(t_1)R(t_3-t_2)$$

$$- R(t_2)R(t_3-t_1) - R(t_3)R(t_2-t_1) \qquad (2.12)$$

While we cannot directly estimate the fourth order cumulant, we can estimate all terms on the RHS of equation (2.12). Also, a close look at equation (2.11) reveals that we can choose a subset of the t_1, t_2, t_3 time lag combinations such that the cumulants are proportional to the individual model coefficients, as we did with the bicovariances in equation (2.3). In general, for an MA(q) model we have (from equation 2.11):

$$c(\ 0,\ 0,-q) = (\mu_4 - 3\sigma^4)(\alpha_0\alpha_q)\alpha_q^2$$

$$c(-1,-1,-q) = (\mu_4 - 3\sigma^4)(\alpha_0\alpha_q)\alpha_{q-1}^2$$

$$c(-2,-2,-q) = (\mu_4 - 3\sigma^4)(\alpha_0\alpha_q)\alpha_{q-2}^2$$

$$\vdots$$

$$c(-q,-q,-q) = (\mu_4 - 3\sigma^4)(\alpha_0\alpha_q)\alpha_o^2 \qquad (2.13)$$

Each of the above cumulants corresponds to a particular combination of time lags and is proportional to the square of one of the coefficients of the moving average process. Therefore, as with the bicovariances, we may use the estimated cumulants (from equation 2.12) to identify model coefficients.

As an example of the use of fourth order cumulants consider estimating c(0,0,-2) for an MA(2) model. This cumulant is proportional to the coefficient α_2^2. From equation (2.12) we obtain:

$$c(0,0,-2) = Tr(0,0,-2) - 3R(0)R(-2)$$
$$= \Sigma\ X_t^3 X_{t-2} - 3(\Sigma\ X_t^2)(\Sigma\ X_t X_{t-2}).$$

Similarly, for c(-1,-1,-2), which is proportional to α_1^2:

$$c(-1,-1,-2) = Tr(-1,-1,-2) - 2R(-1)R(-1) - R(0)R(-2)$$
$$= \Sigma\ X_t X_{t-1}^2 X_{t-2} - 2(\Sigma X_t X_{t-1})^2 - (\Sigma X_t^2)(\Sigma X_t X_{t-2});$$

and finally, for the cumulant proportional to α_0^2, we get:

$$c(-2,-2,-2) = Tr(-2,-2,-2) - 3R(0)R(-2)$$
$$= \Sigma\ X_t X_{t-2}^3 - 3(\Sigma X_t^2)(\Sigma X_t X_{t-2})$$

We now prove the necessity of the condition of non-normality. We have already shown that the alternative MA(q) models are not identifiable using the first two moments, or cumulants, so that moments, or cumulants, beyond the second are needed. Since all the cumulants for the Normal distribution are identically zero beyond the second, the proof of necessity follows.

To complete the theorem all that is needed is to demonstrate that an MA(1) process is identifiable. The induction argument just developed can then be applied to complete the proof that only non-normal models with finite moments can be identified. This part of the proof will also provide an example of the technique.

Let us define two processes:

$$X_t = \epsilon_t - \alpha\epsilon_{t-1}; \qquad |\alpha|<1 \tag{2.14}$$
$$Y_t = \epsilon_t - \beta\epsilon_{t-1}; \qquad \beta = \alpha^{-1}.$$

The $\{\epsilon_t\}$ are assumed to be independently and identically distributed with moments $\mu_2=\sigma_\epsilon^2$ and $\mu_3(\epsilon)$. Using (2.2) it is easy to verify that:

$$E\{X_t^2 X_{t-1}\} = -\alpha\mu_3$$
$$E\{X_t^3\} = \mu_3(1-\alpha^3);$$
$$E\{Y_t^2 Y_{t-1}\} = -\beta\mu_3 \tag{2.15}$$
$$E\{Y_t^3\} = \mu_3(1-\beta^3);$$

and if we now look at the respective ratios to eliminate the presence of μ_3, we obtain:

$$E\{X_t^2 X_{t-1}\}/E\{X_t^3\} = -\alpha/(1-\alpha^3);$$
$$E\{Y_t^2 Y_{t-1}\}/E\{Y_t^3\} = -\beta/(1-\beta^3); \tag{2.16}$$

and identification is achieved if the two expressions in (2.16) cannot be transformed into each other. Recalling that $\beta = \alpha^{-1}$, it is immediately clear that such a transformation is not possible. Consequently, the MA(1) model is identifiable and therefore by induction all MA(q) models are identifiable under the stated conditions. A similar argument using the tricovariances when $\mu_3(\epsilon)=0$ can be used to show that an MA(1) model is identified and therefore all MA(q) models that are non-Gaussian with finite fourth moments.

In what follows we will concentrate mainly on MA(2) processes for expository purposes. However, extensions to higher order processes are straightforward. As noted, for any MA(q) model there exist 2^q possible specifications compatible with the same autocorrelation function, one of which is invertible; consequently, an MA(2) will have four specifications, one of them being invertible. The standard estimation methods available, such as OLS or Box and Jenkins ARIMA modeling, are based on the autocorrelation function and are thus incapable of discriminating among these 2^q alternatives. The implicit assumptions made by these techniques yield only the invertible specification. To find the other 2^q-1 specifications we take this invertible specification, find its roots from equation (1.2), then take different combinations of these roots and their inverses and substitute them into equation (1.3) thus obtaining the complete set of alternative models. Once this is done, we use the bicovariances estimated in equation (2.3) to identify the correct model, i.e., the one compatible with the observed data. (If the third order moment of the innovations is zero we use instead the tricovariances defined in equation 2.12).

To discuss the probability limits of our estimates, we begin by assuming that the estimates of the coefficients $\{\alpha_i\}$ of the MA process obtained by traditional methods have probability limits that equal the MA parameters being estimated. From Slutsky's theorem (see Judge et al, p.147,150), if g(.) is a continuous vector function, then, when plim $\hat{\alpha}_i=\alpha_i$:

$$\text{plim}[g(\hat{\alpha}_i)] = g[\text{plim}(\hat{\alpha}_i)] = g(\alpha_i), \qquad (2.17)$$

where plim $\hat{\alpha}_i=\alpha_i$ is defined by: $\lim_{n\to\infty} \Pr\{|\hat{\alpha}_i-\alpha_i|>\epsilon\}=0$.
Equation (2.17) implies that the roots $\{\hat{\lambda}_i\}$ of the estimated auxiliary equation (1.2) converge in probability to λ_i and that, likewise, the other 2^q-1 sets of estimates obtained by combining these roots in the manner of equation (1.3) also converge in probability to their respective limits. Illustrating this for an MA(2) process, we may establish

$$\text{plim}\ \hat{\lambda}_i = \text{plim}[(-\hat{\alpha}_1 \pm \surd(\hat{\alpha}_1^2 - 4\hat{\alpha}_2))/2\hat{\alpha}_2]$$

$$= (-\text{plim}\ \hat{\alpha}_1 \pm \surd((\text{plim}\ \hat{\alpha}_1)^2 - 4\text{plim}\ \hat{\alpha}_2))/2\text{plim}\ \hat{\alpha}_2$$

from which we conclude that λ_1 and λ_2, and their combinations, converge in probability.

Given that the bicovariances are proportional to the individual coefficients of the correct model, a visual comparison is sometimes sufficient to carry out the selection, or at least, to discard some of the alternative specifications. A more formal, though disarmingly simple, test for identification can be proposed by finding the correlation coefficient, ρ (one for each of the 2^q models), between the series formed by taking the magnitudes of the bicovariances

$$|Bc(-q,-q)|, |Bc(-q+1,-q)|, \ldots, |Bc(0,-q)| \qquad (2.18)$$

and the series formed by taking the magnitudes of the coefficients to which they are proportional:

$$|\alpha_0|, |\alpha_1|, \ldots, |\alpha_q|. \qquad (2.19)$$

The model with the highest positive correlation is selected. The same type of test may be used with the cumulants of equation (2.12). This test can be explained as follows. Our problem, once we have the 2^q sets of alternative coefficients, is to choose the correct one. We know that the set of bicovariances in equation (2.18) is proportional to the corresponding sequence of the absolute values of the coefficients of the correct model. Therefore, by searching for the largest positive correlation between the bicovariances in (2.18) and each alternative set of coefficients we are, in effect, looking for that set of absolute values of the coefficients $\{\alpha_i\}$ which exhibits the "best fit" to the bicovariances. We can write in general:

$$|Bc_i(\cdot)| = k_i |\alpha_i|$$

where $k_i>0$, $i=1,\ldots,q$, is a factor of proportionality between the magnitudes of the bicovariance and the corresponding model coefficient. Therefore, letting $\hat{\alpha}_i$ denote the estimates obtained by the procedure outlined above, we have:

$$\rho = \mathrm{cov}(|\hat{B}c_i|, |\hat{\alpha}_i|)/\surd\mathrm{var}(|\hat{B}c_i|)\surd\mathrm{var}(|\hat{\alpha}_i|)$$

$$= \mathrm{cov}(k_i|\hat{\alpha}_i|, |\hat{\alpha}_i|)/\surd\mathrm{var}(k_i|\hat{\alpha}_i|)\surd\mathrm{var}(|\hat{\alpha}_i|)$$

$$= (E\ k_i\hat{\alpha}_i^2 - E\ |\hat{\alpha}_i|\ E\ k_i|\hat{\alpha}_i|)/\surd\mathrm{var}(k_i|\hat{\alpha}_i|)\surd\mathrm{var}(|\hat{\alpha}_i)$$

(2.20)

From the above equation we can see that only if, for all i, $k_i=k$, where k is some constant, will $\rho=1$ since we are then able to pull k_i out of the expectation operator and write

$$\rho = k\ \mathrm{var}(|\hat{\alpha}_i|)/k\ \mathrm{var}(|\hat{\alpha}_i|) = 1$$

The condition that all k_i's be equal for a given coefficient set implies the true model which in turn implies a perfect fit with the bicovariances.

Returning to the MA(2) example, the bicovariance series is

$$|Bc(-2,-2)|, |Bc(-1,-2)|, |Bc(0,-2)|$$

which would be correlated against the set of absolute values of the coefficients

$$|\alpha_0|, |\alpha_1|, |\alpha_2|$$

from each one of the four alternative models.

To illustrate the identification procedure we generate a sequence $\{x_t\}$ using equation (1.8a) from our previous example and repeated here for convenience,

$$x_t = \epsilon_t - 2.33\epsilon_{t-1} + .666\epsilon_{t-2} \tag{1.8a}$$

with roots $\{\lambda_i\}=(2,1/3)$.

The sequence $\{\epsilon_t\}$ used to generate $\{x_t\}$ consisted of 1000 draws from a zero mean, variance one, exponential distribution with skewness, μ_3, equal to 2.2.

Suppose we observe only the sequence $\{x_t\}$ and are asked to estimate the parameters of the underlying model. We begin by applying the Box and Jenkins ARIMA procedure to the observations $\{x_t\}$ to obtain (t values in parentheses),

$$\hat{x}_t = \epsilon - \underset{(-26.1)}{.8233}\epsilon_{t-1} + \underset{(4.3)}{.1366}\epsilon_{t-2}, \tag{2.21}$$

as an estimate of the invertible model shown in equation (1.8c) with roots (1/2,1/3). Using equation (1.2) we can re-express the estimated model as

$$\hat{x}_t = (1 - .8233B + .1366B^2)\epsilon_t \tag{2.22}$$

The estimated roots of the auxiliary equation above are 4.35 and 1.685. The set of autocorrelation equivalent specifications are found by substituting different combinations of the roots, and their inverses, into equation (2.1) to obtain:

	α_0	α_1	α_2
model 1	1	$-(4.35+1.685) = -6.00$	$(4.35)(1.685) = 7.33$
model 2	1	$-(4.35+1.685^{-1}) = -4.90$	$4.35/1.685 = 2.58$
model 3	1	$-(1.685+4.35^{-1}) = -1.90$	$1.685/4.35 = 0.39$
model 4	1	$-(4.35^{-1}+1.685^{-1}) = -0.82$	$\{(4.35)(1.685)\}^{-1} = 0.14$

The estimates of the bicovariances for these data yield:

$$|Bc(-2,-2)| = |\Sigma x_t x_{t-2} x_{t-2}| = 2550 \qquad \propto|\alpha_0|$$

$$|Bc(-1,-2)| = |\Sigma x_t x_{t-1} x_{t-2}| = 4466 \qquad \propto|\alpha_1|$$

$$|Bc(0,-2)| = |\Sigma x_t x_t x_{t-2}| = 601 \qquad \propto|\alpha_2|,$$

for the correct model.

The correlation coefficients, ρ, between the bicovariance series {2550,4466,601} and the magnitudes of the coefficients is estimated for each model. These correlations were -.2, .59, .99, .756 respectively for model 1 to 4. On this basis, model 3:

$$x_t = \epsilon_t - 1.9\epsilon_{t-1} + .39\epsilon_{t-2} \qquad (2.23)$$

is correctly chosen as the estimate of the process we generated using equation (1.8a). In this example the fact that the .99 correlation was easily singled out might leave an overly optimistic impression about the effectiveness of the method. Further, the above results indicate that the estimation of the roots is sensitive to the estimation of the invertible coefficients. In short, it would appear that the mean square error of estimate for the non-invertible model's coefficients are substantially greater than the mean square error of the invertible model's coefficients.

(3) IMPLICATIONS OF IDENTIFIABILITY

A) Recovery of the Innovations

In other disciplines, seismology for instance, there is great interest in estimating the innovations $\{\epsilon_t\}$ due to their information content about the underlying physical system. Recovery of the sequence $\{\epsilon_t\}$, known as deconvolution, presupposes that the identification and estimation stages have been carried out, i.e., the coefficients $\{\alpha_i\}$ have been estimated. Once this is achieved and the roots of the auxiliary equation (1.2) found, we can easily solve for $\hat{\epsilon}_t$, our estimate of ϵ_t, by solving equation (1.3). The resultant expression for $\hat{\epsilon}_t$ may be a function of past and present x's, of future x's, or a mixture of past and future x's. The $\{\hat{\epsilon}_t\}$ recovered by traditional methods, say by taking the OLS residuals or through Box and Jenkins ARIMA modeling, yield consistent estimates of $\{\epsilon_t\}$ only if the observations come from an invertible model.

It should be noted that some innovations at the beginning or at the end of the sequence $\{\epsilon_t\}$ can not be estimated. In the invertible case, for example, the expression for $\hat{\epsilon}_t$ is of the general form

$$\hat{\epsilon}_t = \hat{\psi}_0 x_t + \hat{\psi}_1 x_{t-1} + \hat{\psi}_2 x_{t-2} + \dots \quad (3.1)$$

where $\{\hat{\psi}_s\}$ is a decreasing sequence of estimated coefficients obtained by solving for $\hat{\epsilon}_t$ in terms of $\{\hat{\alpha}_i\}$ and the observed $\{x_t\}$. It is clear from this expression that the innovations at the beginning of the sequence cannot be recovered due to the lack of previous observations on x_t. In the strictly non-invertible case, the expression for $\hat{\epsilon}_t$ is of the form

$$\hat{\epsilon}_t = \hat{\xi}_1 x_{t+1} + \hat{\xi}_2 x_{t+2} + \hat{\xi}_3 x_{t+3} + \dots \quad (3.2)$$

where $\{\hat{\xi}_s\}$ forms a decreasing sequence in time. If, say, we have observations up to time T, the innovations close to, and at T, cannot be estimated because we do not have future observations on x_t, i.e., for t>T. This fact has important negative implications for our ability to improve forecasts by identifying the correct model. If we were allowed to retrieve ϵ_T, then a linear forecast for x_{T+1} would be superior in the non-invertible case compared to that of the invertible case, by virtue of the higher weight assigned to ϵ_t in the former model; a discussion about the irrelevancy of invertibility to linear forecasting is found in Granger and Newbold (1977, p.144). However, while this is the case for linear methods, the door is still slightly open for research aiming at improved forecasts via non-linear methods.

To illustrate the recovery of the innovations we recall the estimated equation (2.23) and solve it for $\hat{\epsilon}_t$ as follows

$$x_t = \hat{\epsilon}_t - 1.9\hat{\epsilon}_{t-1} + .39\hat{\epsilon}_{t-2},$$

or $$x_t = (1 - 1.9B + .39B^2)\hat{\epsilon}_t,$$

or $$x_t = .39(B - 4.35)(B - .59)\hat{\epsilon}_t,$$

or $$x_t = -1.69(1 - .23B)(1 - .59B^{-1})B\hat{\epsilon}_t,$$

in order to obtain a convergent expansion in terms of the sequence $\{x_t\}$. After suitable truncation we obtain:

$$\hat{\epsilon}_t = -.03x_{t-1} - .15x_t - .68x_{t+1} - .405x_{t+2} - .24x_{t+3} \tag{3.3}$$

Figure 1 compares a section of the original exponential innovations ϵ_t with our estimate $\hat{\epsilon}_t$ from the above equation. Figure 2 compares the original innovations with those that would be obtained by incorrectly assuming invertibility and running OLS on the observations; the OLS estimates of ϵ_t are the residuals of the regression equation:

$$x_t = -.813x_{t-1} - .512x_{t-2} - .33x_{t-3} - .14x_{t-4} \qquad (3.4)$$
$$\quad (-26) \qquad (-13) \qquad (-8)$$

where the t values are shown in parentheses.

B) The Generation of ARCH Results

This section shows how attempts to forecast by running OLS using observations from a noninvertible MA process result in residuals with ARCH structure. In contrast, no ARCH effects are generated if the process is invertible. For simplicity, moving averages of order one will be used.

Consider the following invertible (3.5a) and noninvertible (3.5b) models, appropriately labeled xi for invertible and xn for noninvertible:

$$xi_t = \epsilon_t + \alpha\epsilon_{t-1} \qquad |\alpha|<1 \qquad (3.5a)$$

$$xn_t = \alpha\epsilon_t + \epsilon_{t-1} \qquad |\alpha|<1 \qquad (3.5b)$$

Figure 1

ORIGINAL VS OUR ESTIMATED INNOVATIONS

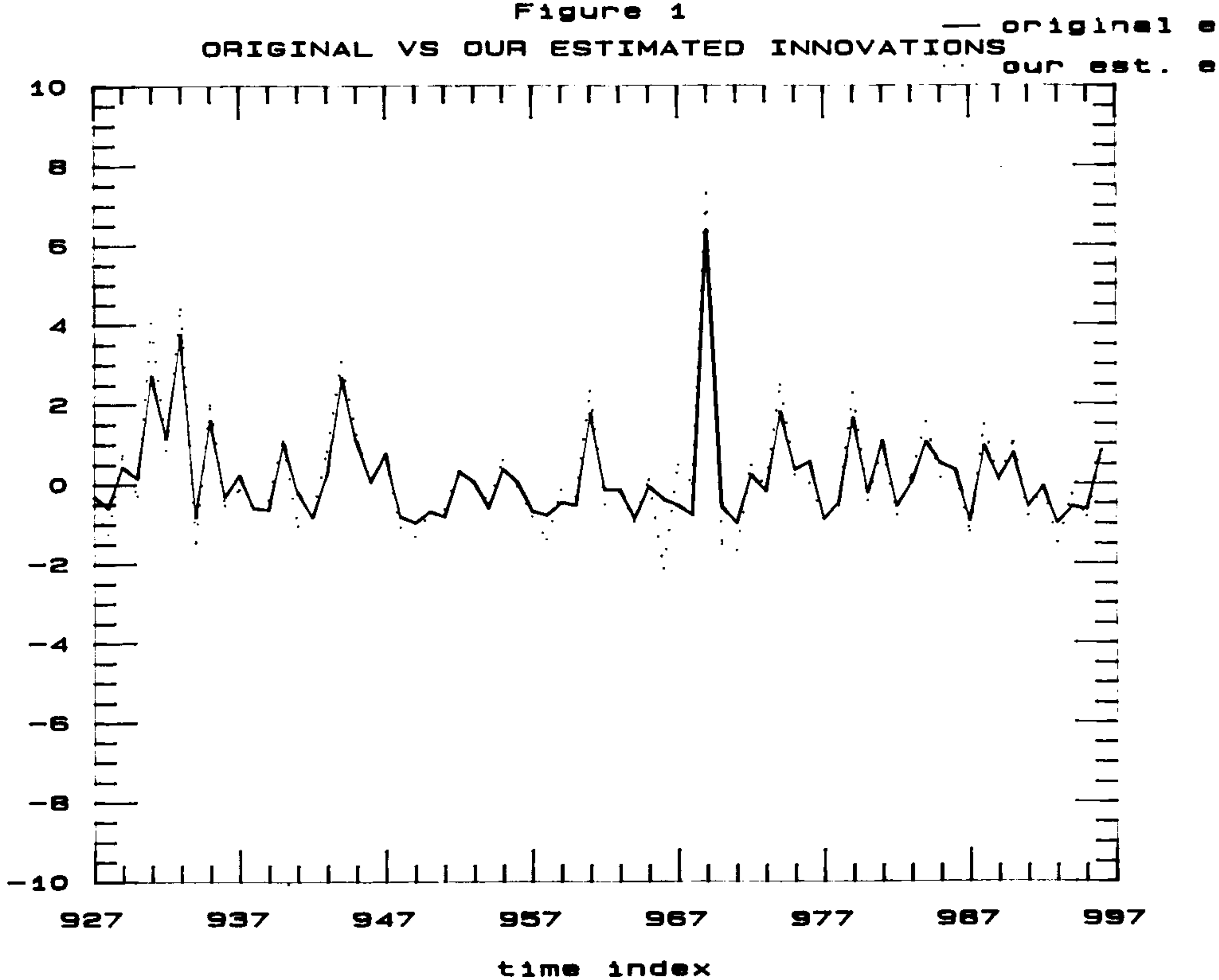
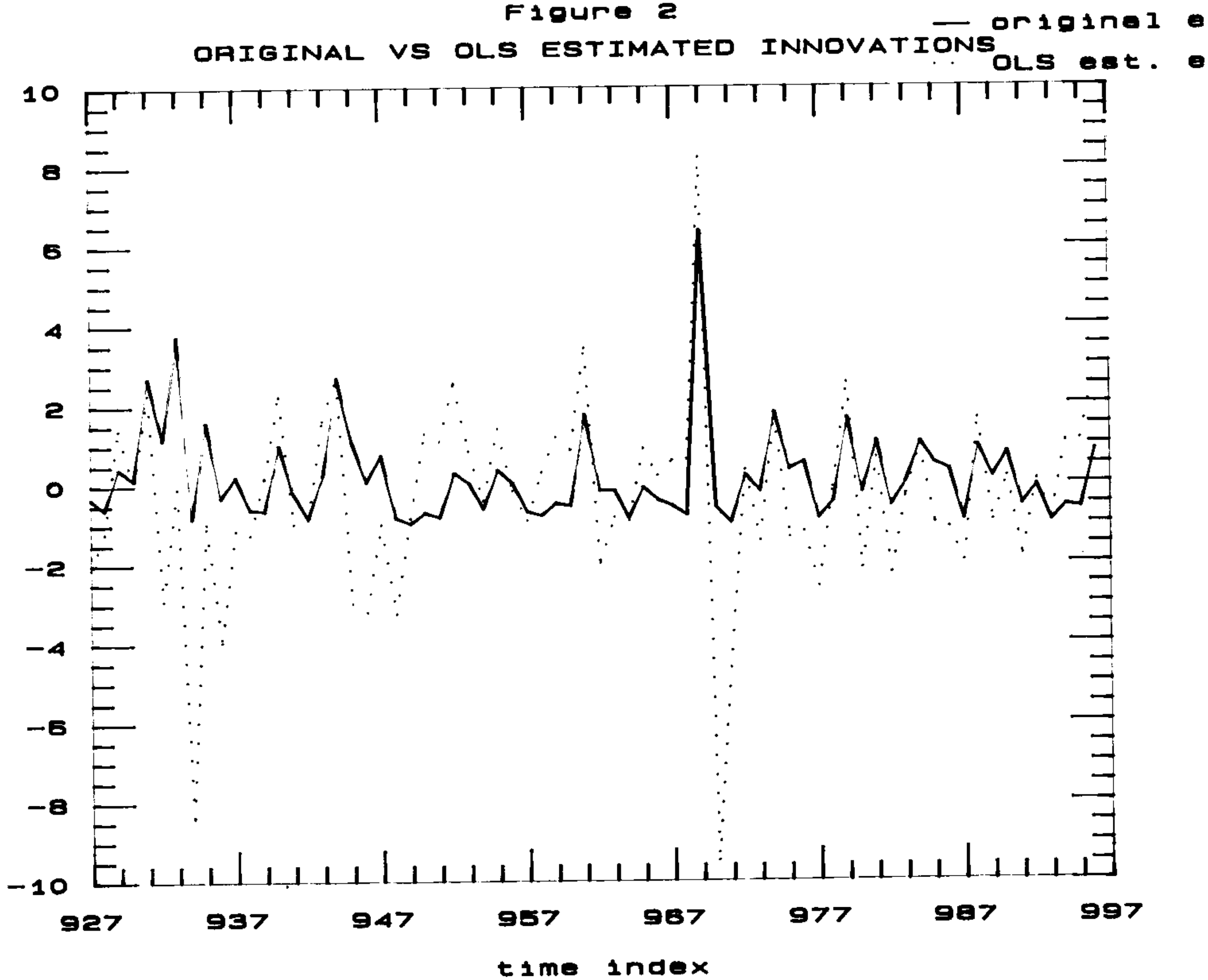


Figure 2

ORIGINAL VS OLS ESTIMATED INNOVATIONS



where the coefficient α is the same in both equations.

The one-step-ahead forecast for the invertible case, denoted here as $\hat{xi}_t$, is found by running OLS to obtain

$$\hat{xi}_t = \hat{\alpha} xi_{t-1} - \hat{\alpha}^2 xi_{t-2} + \hat{\alpha}^3 xi_{t-3} - \dots \tag{3.6}$$

which is simply the inverted (or autoregressive) representation of the MA process.

Given that OLS is based on the autocorrelation properties of the process, we obtain the same result in the noninvertible case:

$$\hat{xn}_t = \hat{\alpha} xn_{t-1} - \hat{\alpha}^2 xn_{t-2} + \hat{\alpha}^3 xn_{t-3} - \dots \tag{3.7}$$

While the estimated coefficients of both equations are the same, their OLS residuals differ. To facilitate comparisons let us substitute for the estimates, their probability limits. Substituting model (3.5a) into equation (3.6) we get

$$\begin{aligned} \tilde{xi}_t &= \alpha[\epsilon_{t-1} + \alpha\epsilon_{t-2}] - \alpha^2[\epsilon_{t-2} + \alpha\epsilon_{t-3}] + \alpha^3[\epsilon_{t-3} + \alpha\epsilon_{t-4}] - \dots \\ &= \alpha\epsilon_{t-1} \end{aligned} \tag{3.8}$$

where all terms other than $\alpha\epsilon_{t-1}$ cancel. The OLS residual for the invertible case, labeled Ir, is therefore

$$Ir = xi_t - \tilde{x}i_t = \epsilon_t + \alpha\epsilon_{t-1} - \tilde{x}i_t$$

$$= \epsilon_t + \alpha\epsilon_{t-1} - \alpha\epsilon_{t-1} = \epsilon_t \qquad (3.9)$$

In contrast, the residuals defined with respect to the finite sample size estimate $\hat{\alpha}$ are:

$$\epsilon_t + \delta\epsilon_{t-1} - \hat{\alpha}\delta\epsilon_{t-2} + \hat{\alpha}^2\delta\epsilon_{t-3} - \hat{\alpha}^3\delta\epsilon_{t-4} + \ldots \qquad (3.10)$$

where $\delta=\alpha-\hat{\alpha}$. Because $\hat{\alpha}$ has a probability limit α, $|\alpha|<1$, the expression in equation (3.10) converges rapidly in probability to ϵ_t. Even for finite samples, one would only need the first few terms to obtain a very close approximation. We define these residuals by $\hat{I}r$.

For the noninvertible case we substitute model (3.5b) into the regression in equation (3.7) to obtain a different forecast in which the lag innovations do not cancel

$$\tilde{x}n_t = \alpha[\alpha\epsilon_{t-1} + \epsilon_{t-2}] - \alpha^2[\alpha\epsilon_{t-2} + \epsilon_{t-3}] + \alpha^3[\alpha\epsilon_{t-3} + \epsilon_{t-4}] - \ldots$$

$$= \alpha^2\epsilon_{t-1} + (\alpha-\alpha^3)\epsilon_{t-2} - (\alpha^2-\alpha^4)\epsilon_{t-3} + (\alpha^3-\alpha^5)\epsilon_{t-4} - \ldots \qquad (3.11)$$

Finally, the OLS error for the noninvertible case, Nr, is

$$Nr = xn_t - \tilde{x}n_t = \alpha\epsilon_t + \epsilon_{t-1} - \tilde{x}n_t$$

$$= \alpha\epsilon_t + (1-\alpha^2)\epsilon_{t-1} - (\alpha-\alpha^3)\epsilon_{t-2} + (\alpha^2-\alpha^4)\epsilon_{t-3} - \ldots \qquad (3.12)$$

As with the invertible case, the substitution of $\hat{x}n_t$ for $\tilde{x}n_t$ yields a rapidly converging sequence in powers of $\hat{\alpha}$ and δ associated with lagged values of ϵ_t.

Notwithstanding the differences in residuals exhibited in equations (3.9) and (3.12), the mean square prediction error, MSPE, is the same in both cases. For the invertible case, using equation (3.9), we find the MSPE to be

$$\text{MSPE} = E[Ir]^2 = E[(xi_t - \tilde{x}i_t)^2] = E(\epsilon_t)^2 = \sigma_\epsilon^2 \qquad (3.13)$$

For the noninvertible case we use equation (3.12) and note that terms containing cross products of the $\{\epsilon_t\}$ will vanish once the expectation is taken:

$$\text{MSPE} = E[Nr]^2 = E[(xn_t - \tilde{x}n_t)^2] =$$

$$= E[\alpha^2\epsilon^2 + (1-\alpha^2)^2\epsilon_{t-1}^2 + (\alpha-\alpha^3)^2\epsilon_{t-2}^2 + (\alpha^2-\alpha^4)^2\epsilon_{t-3}^2 + (\alpha^3-\alpha^5)^2\epsilon_{t-4}^2 + \ldots]$$

$$= \sigma_\epsilon^2[\alpha^2 + 1 - 2\alpha^2 + \alpha^4 + \alpha^2 - 2\alpha^4 + \alpha^6 + \alpha^4 - 2\alpha^6 + \alpha^8 + \alpha^6 - 2\alpha^8 + \alpha^{10} + \alpha^8 - 2\alpha^{10} + \alpha^{12} + \ldots]$$

$$= \sigma_\epsilon^2 \qquad (3.14)$$

The usual test for ARCH structure, variances that are themselves conditionally autoregressive, is to regress the OLS residuals squared on past x's; significant asymptotic t values for these variables is indicative of an ARCH structure; see Engle (1982). Estimating the residuals from a non-invertible model when invertibility is assumed produces an ARCH structure in the squared residuals. To see this, square the noninvertible OLS residuals and regress them on xn_{t-1}, to obtain:

$$\hat{b} = \text{plim } \{[\Sigma \hat{N}r_t^2 \; xn_{t-1}]/[\Sigma \; xn_{t-1}^2]\}, \tag{3.15}$$

where $\hat{b}$ is the estimate of the linear relationship between Nr_t^2 and xn_{t-1}. Substituting (3.12) and (3.5b) into (3.15) and noting that terms with cross ϵ_t's will vanish in the probability limit, we get

$$\text{plim } \hat{b} =$$

$$\text{plim } \{\Sigma[\hat{\alpha}\epsilon_t+(1-\hat{\alpha}^2)\epsilon_{t-1}-(\hat{\alpha}-\hat{\alpha}^3)\epsilon_{t-2}+$$

$$(\hat{\alpha}^2-\hat{\alpha}^4)\epsilon_{t-3}-\ldots)^2 \; (\hat{\alpha}\epsilon_{t-1}+\epsilon_{t-2})]/\Sigma[\hat{\alpha}\epsilon_{t-1}+\epsilon_{t-2}]^2\}$$

$$= \text{plim } \{\Sigma[\hat{\alpha}(1-\hat{\alpha}^2)^2\epsilon_{t-1}^3 + (\hat{\alpha}-\hat{\alpha}^3)^2\epsilon_{t-2}^3]/\Sigma[\hat{\alpha}\epsilon_{t-1} + \epsilon_{t-2}]\}^2$$

$$= g(\alpha)\mu_3(\epsilon)/\sigma_\epsilon^2, \tag{3.17}$$

where g is a function of α.

In this example we obtain significant ARCH effects as long as the third order moment of the innovations is nonzero. When the innovations have a zero third moment, such as with a normal distribution, no ARCH effects would be found, whether invertible or not.

To demonstrate the appearance of ARCH effects we have generated 4 pairs of models, where the first model in each pair is invertible and the second is noninvertible. All sequences are 1000 observations long and are described as follows:

pair 1: $yi_t = u_t + .2u_{t-1}$

$yn_t = .2u_t + u_{t-1}$

where u_t is distributed as exponential with variance of 1 and μ_3=1.6.

pair 2: $xi_t = 6\epsilon_t + 5\epsilon_{t-1} + \epsilon_{t-2}$

$xn_t = \epsilon_t + 5\epsilon_{t-1} + 6\epsilon_{t-2}$

where ϵ_t is distributed as exponential with variance of 1 and μ_3=2.2.

pair 3: $zi_t = 6v_t + 5v_{t-1} + v_{t-2}$

$zn_t = v_t + 5v_{t-1} + 6v_{t-2}$

where v_t is distributed N(0,1)

pair 4: $si_t = 6d_t + 5d_{t-1} + d_{t-2}$

$sn_t = d_t + 5d_{t-1} + 6d_{t-2}$

where d_t is a uniform distribution with zero mean and a variance of 1.

OLS was used to estimate the invertible version of all eight models. The next step was to square the residuals and regress them on various lags of the observations plus a constant. The results are shown by pairs below. Notice that the t values for the lagged variables are higher in the noninvertible cases than in the invertible cases except, as expected, when the innovations are normal or symmetrically distributed. Likewise, the F values for the non-invertible models in the first two pairs are significant at the 1% level, but not significant with the unidentifiable non-invertible models in pairs 3 and 4.

pair 1: $(yiRES)^2 = .89 - .06yi_{t-1} + .12yi_{t-2} - .04yi_{t-3}$

$(14.9) \quad (-.9) \quad (1.9) \quad (-.7) \qquad F_{3,993}=1.37$

$(ynRES)^2 = .9 + .28yn_{t-1} + .07yn_{t-2} - .05yn_{t-3}$

$(15.9) \quad (4.7) \quad (1.2) \quad (-.8) \qquad F_{3,993}=8.74$

pair 2: $(xiRES)^2 = 40 - .2xi_{t-1} - .13xi_{t-2} - .5xi_{t-3}$

$(10.5) \quad (-.3) \quad (-.2) \quad (-.9) \qquad F_{3,992}=1.00$

$(xnRES)^2 = 39 + 2.9xn_{t-1} + .13xn_{t-2} + .8xn_{t-3}$

$(16.7) \quad (8.2) \quad (.3) \quad (2.1) \qquad F_{3,992}=42.9$

pair 3: $(ziRES)^2 = 37 - .24zi_{t-1} + .3zi_{t-2} - .5zi_{t-3}$

$(22.5) \quad (-.9) \quad (.9) \quad (-1.8) \qquad F_{3,991}=1.27$

$$(znRES)^2 = 37 + .14zn_{t-1} - .02zn_{t-2} - .4zn_{t-3}$$
$$(21.4) \quad (.5) \qquad (-.1) \quad (-1.3) \qquad F_{3,991}=1.07$$

pair 4: $$(siRES)^2 = 38 - .15si_{t-1} + .42si_{t-2} - .1si_{t-3}$$
$$(34.8) \quad (-.9) \qquad (2.0) \quad (-.6) \qquad F_{3,991}=1.68$$

$$(snRES)^2 = 38 + .06sn_{t-1} + .14sn_{t-2} + .07sn_{t-3}$$
$$(24.4) \quad (.2) \qquad (.5) \quad (.3) \qquad F_{3,991}=.47$$

The critical F values for $F_{3,\infty}$ are:

3.78(1% level), 2.60(5% level), 2.08(10% level), 1.37(25% level)

(4) SOME EMPIRICAL EVIDENCE OF NONINVERTIBILITY

In this section we examine the practical relevance of our procedures. This is an important exercise in that it may well be true that all economic time series that can be represented as moving average processes are in fact invertible. Further, the presence of extra noise, the inevitable approximations that are involved, not to mention the very limited number of observations available in most economic time series, may vitiate our results.

The data base that we used to search for time series that might prove amenable to our analysis was CITIBASE of Citibank. The time series that we considered, at least briefly, included: monthly M1, currency, demand deposits and the monetary base; income velocity, quarterly GNP and the GNP deflator;

monthly prime rate and the LIBOR rate; monthly exchange rates for Canada, Japan, United Kingdom and W. Germany with respect to the US dollar; expenditure for new plant and equipment; new construction put in place; monthly unemployment rate; some stock market data; and some leading indicators.

All economic time series that were scanned in any detail were transformed into percentage changes of the original series; we found this procedure to yield generally acceptable stationarity in the series while still preserving their economic significance. Our interest, at this preliminary stage, was to find time series that could be modelled as very low order moving averages, MA(1) or MA(2). While non-zero third moments were commonly estimated, the invertibility assumption was supported in most of the series that we scanned; a finding that agrees with the conventional wisdom. However, we can present some evidence of noninvertibilty in the series for the percentage change in the Prime Rate and percentage change in Expenditure for New Plant and Equipment.

The Prime Rate series includes monthly averages of daily figures, in terms of percent changes (mean subtracted). 200 observations from April 1970 to November 1986: Citibase code FYPR, were observed. The autocorrelation function for the prime rate series is shown in Figure 3; two lags were selected as significant: $r(1)=.59$ and $r(2)=.19$. The estimated autocovariance function is $\hat{R}(0)=31.5$, $\hat{R}(1)=18.6$ and $\hat{R}(2)=6$. The estimated standarized skewness is 2.8.

The Box and Jenkins procedure applied to these observations yields the following invertible specification (t values in parentheses):

ESTIMATED AUTOCORRELATIONS FOR % CHANGE IN THE PRIME RATE

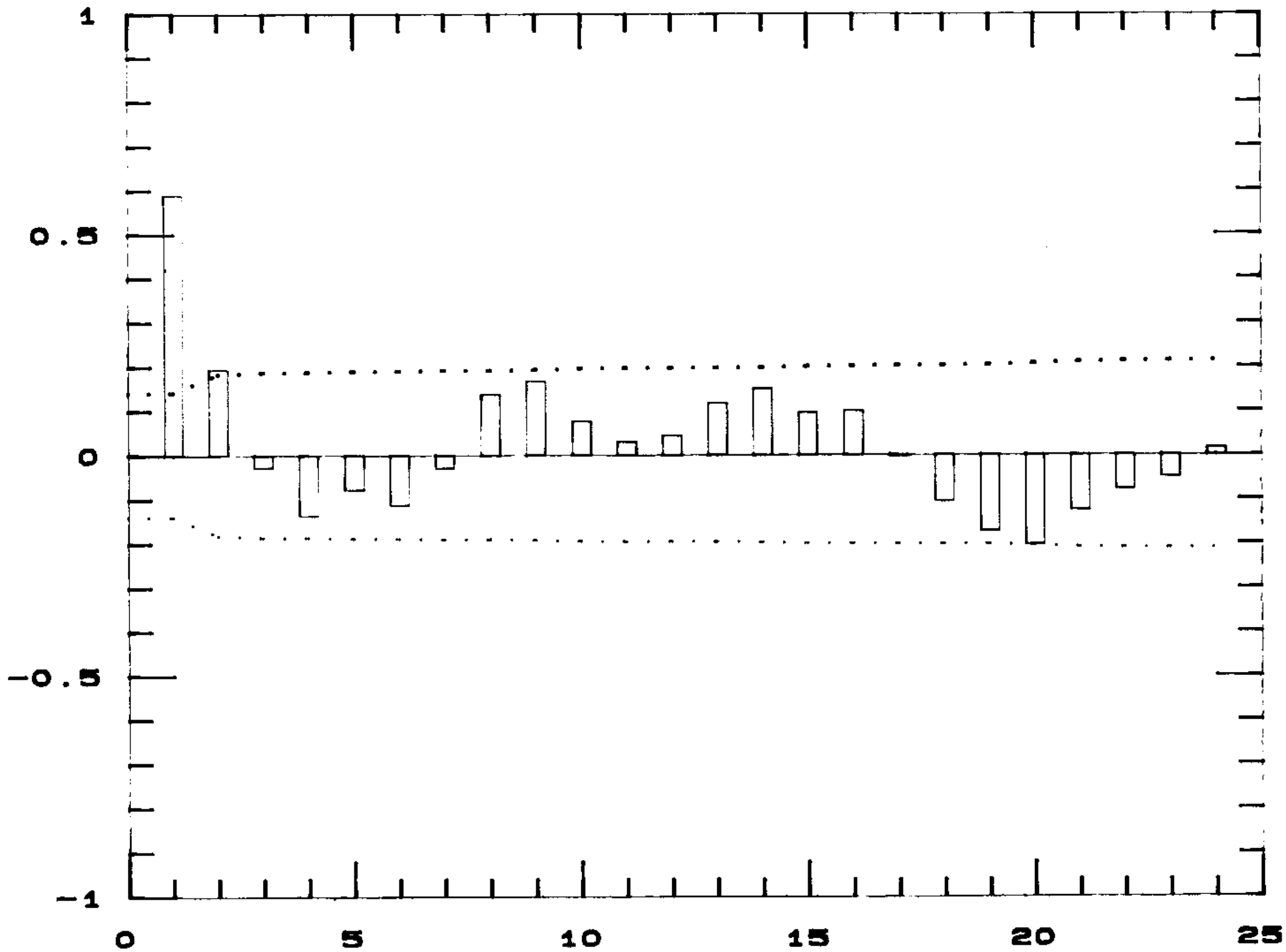


FIGURE 3

$$PR_t = \epsilon_t + .67\epsilon_{t-1} + .19\epsilon_{t-2}. \qquad (4.1)$$
$$\qquad (9.7) \qquad (2.8)$$

The roots of the above equation are -1.73 ± i1.47, where i is $\sqrt{-1}$. Combining these and their inverses, by pairs, we get four autocorrelation equivalent models, of which two will have complex coefficients. Use of these coefficients implies that the innovations would also be complex to assure real valued observations of the prime rate. We assume that the innovations, or shocks to the economy, are real and consequently disregard models with complex coefficients. The two alternative specifications with real coefficients are:

$$PR_t = \epsilon_t + .67\epsilon_{t-1} + .19\epsilon_{t-2}, \text{ invertible;}$$
$$(4.2)$$
$$PR_t = \epsilon_t + 3.46\epsilon_{t-1} + 5.15\epsilon_{t-2}, \text{ non-invertible.}$$

As the reader can appreciate, the two versions differ significantly. The invertible one, obtained by traditional estimation procedures, exhibits descending coefficient values; the noninvertible model has ascending coefficients, that is, the major effect of an innovation is with a lag of 2.

From the data, the estimated bicovariances are:

Bc(-2,-2) = -5568

Bc(-1,-2) = 3330

$Bc(0,-2) = 13829$

The correlations between the absolute values of the bicovariance series and the absolute values of the coefficients for the two models are respectively $\rho_1=-.82$ and $\rho_2=.68$; thus the second alternative, the noninvertible, is selected as the appropriate MA expression for the percentage change in the monthly prime rate:

$$PR_t = \epsilon_t + 3.46\epsilon_{t-1} + 5.15\epsilon_{t-2} \tag{4.3}$$

The sequence of innovations that drives the prime rate process is found by solving for ϵ_t in equation (4.3):

$$\hat{\epsilon}_t = .19PR_{t+2} - .13PR_{t+3} + .05PR_{t+4} - .01PR_{t+5} + \ldots \tag{4.4}$$

A plot of $\{\hat{\epsilon}_t\}$ versus time is shown in Figure 5.

The data on Expenditure for New Plant and Equipment is quarterly percent changes (mean subtracted), for all industries, in current dollars; there are 161 observations from 1947/1 to 1987/2; Citibase code IXI. The estimated autocorrelation function for this series is shown in Figure 4; two lags were selected as significant: r(1)=.52 and r(2)=.31. The estimated autocovariance function is $\hat{R}(0)=10.6$, $\hat{R}(1)=5.5$ and $\hat{R}(2)=3.3$. The estimated standarized skewness is -2.95.

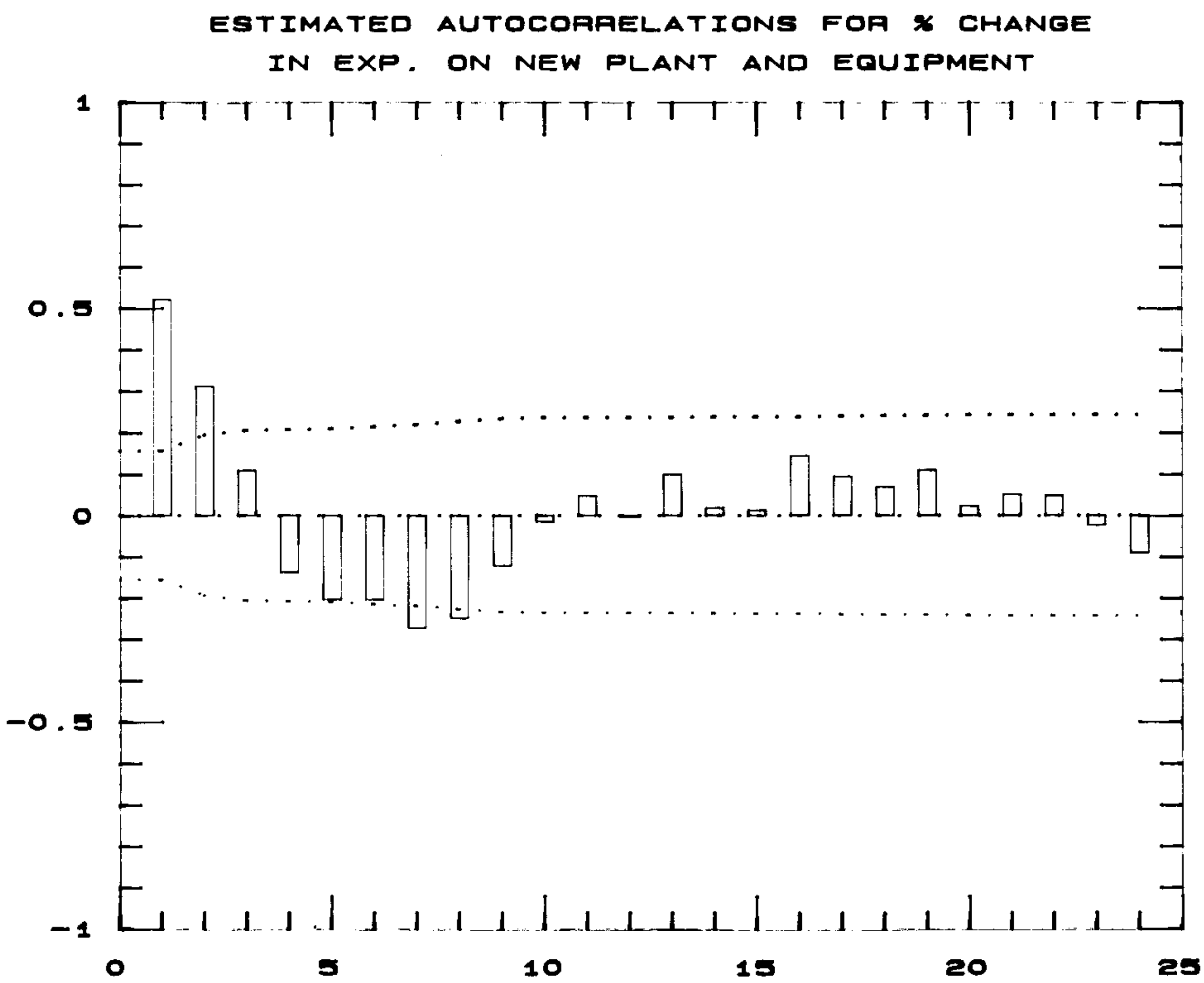
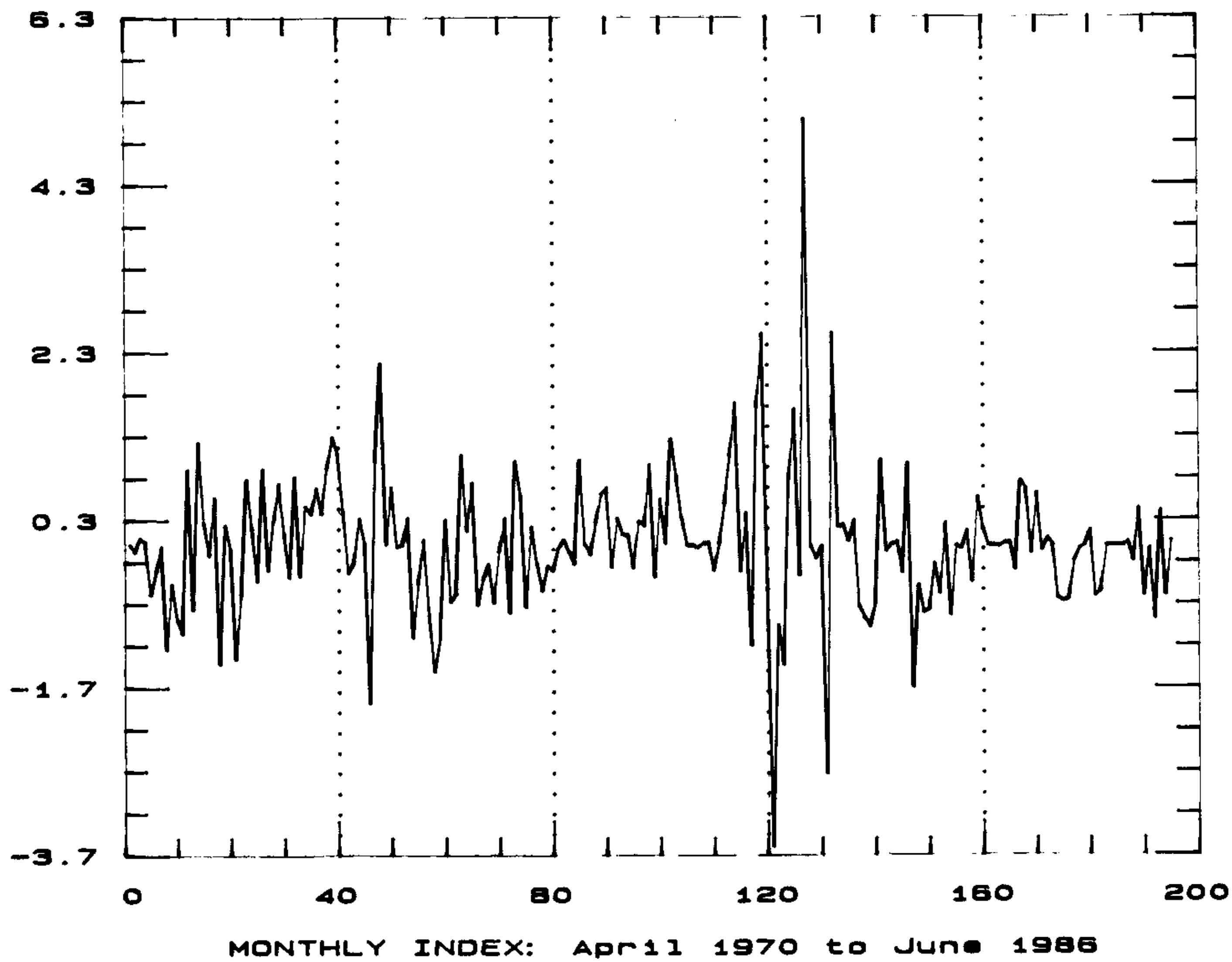

ESTIMATED AUTOCORRELATIONS FOR % CHANGE
IN EXP. ON NEW PLANT AND EQUIPMENT
1
0.5
0
-0.5
-1
0
5
10
15
20
25
lag

FIGURE 4

Figure 5

ESTIMATED INNOVATIONS FOR % CHANGE IN THE PRIME RATE



The Box and Jenkins procedure yields the following estimation (t values in parentheses):

$$NPE_t = u_t + .428u_{t-1} + .23u_{t-2}. \qquad (4.5)$$
$$\qquad\qquad (5.6) \qquad\quad (3.0)$$

The roots are complex and form the conjugate pair $-.93 \pm i1.87$. The two specifications with real valued coefficients are

$$NPE_t = u_t + .428u_{t-1} + .23u_{t-2}, \text{ invertible;}$$
$$(4.6)$$
$$NPE_t = u_t + 1.86u_{t-1} + 4.36u_{t-2}, \text{ non-invertible.}$$

From the data the estimated bicovariances are:

$Bc(-2,-2) = -1027$

$Bc(-1,-2) = -1385$

$Bc(\ 0,-2) = -1572$

The correlations between the absolute values of the bicovariances and the coefficients of the two models are ρ_1=-.996 and ρ_2=.9, suggesting the noninvertible alternative as appropriate, that is,

$$NPE_t = u_t + 1.86u_{t-1} + 4.36u_{t-2} \quad (4.7)$$

The innovations that drive this process are found by solving for $\{u_t\}$ in equation (4.7):

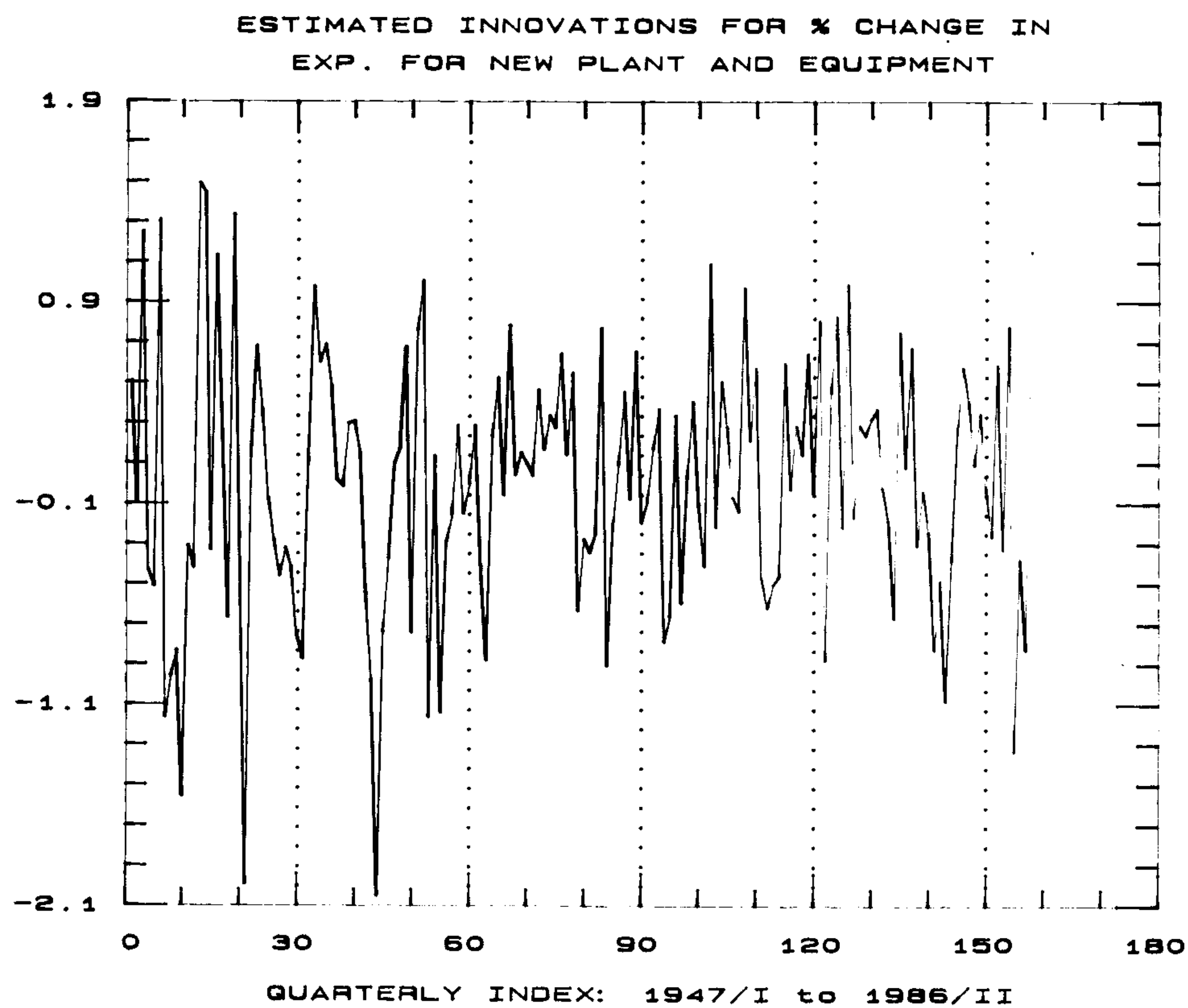
$$\hat{u}_t = .23NPE_{t+2} - .1NPE_{t+3} - .01NPE_{t+4} + \ldots \quad (4.8)$$

A plot of $\{\hat{u}_t\}$ versus time is shown in Figure 6.

(5) DISCUSSION

While we acknowledge the statistical difficulties that accompany such limited samples as those presented in the previous section, the results obtained in these examples have important implications for the way that we model economic behavior in terms of the timing of an agent's reactions to incoming innovations or shocks. Invertibility, the usual assumption in traditional estimation procedures, automatically places restrictions on the value that the moving average parameters can take. It rules out, for example, the case where $|\alpha_1|$ is greater than $|\alpha_0|$ in an MA(1) model. It rules out, in other words, the possibility that economic agents react more vigorously to an innovation ϵ_t after some time has elapsed. Similar statements can be made about higher order moving average processes MA(q), where, if invertibility is imposed, $|\alpha_q|$ can never be greater than $|\alpha_0|$. We must note, however, that no simple ranking rule is available for the other intervening coefficients of the model.

Figure 6



ESTIMATED INNOVATIONS FOR % CHANGE IN
EXP. FOR NEW PLANT AND EQUIPMENT
1.9
0.9
-0.1
-1.1
-2.1
0
30
60
90
120
150
180
QUARTERLY INDEX: 1947/I to 1986/II

Our results tentatively indicate that the variables, prime rate and expenditure on new plant and equipment, both in percent changes, delay their strongest reactions to incoming innovations, by 2 months for the prime rate and by 2 quarters for the expenditure figures.

If forecasting $\{x_t\}$ is the only criterion and if we restrict ourselves to linear forecasts, then non-invertibility is an irrelevant issue. But if one is concerned with the recoverability of the $\{\epsilon_t\}$ and the timing of responses to shocks, then invertibility should be tested in these situations as a matter of course.

A quick way to check for invertibility is to estimate only the bicovariances proportional to the first and last coefficients, that is, from equation (2.3),

$$|Bc(-q,-q)| = |\Sigma x_t x_{t-q} x_{t-q}|, \qquad \text{which is proportional to } |\alpha_0|$$

$$|Bc(\ 0,-q)| = |\Sigma x_t x_t x_{t-q}|, \qquad \text{which is proportional to } |\alpha_q|$$

then, if $|Bc(-q,-q)| < |Bc(0,-q)|$, one should be wary of assuming invertibility.

Thus, if we are interested in the information content of the innovations, then the identification of the correct model is vital. To date, forecastability has been an almost sole criterion for the estimation of an MA process. Perhaps it is time to shift our research emphasis towards a better understanding of the underlying process, even though it might have little initial effect on forecasting. As we have seen the estimated time paths of the invertible and non-invertible innovations are very different. Consequently, we now have the opportunity to begin asking more searching questions about both the statistical properties of the innovations and their informational and economic content. As a modest beginning, we can now attempt to relate the identified innovation time paths to economic events, news releases, such as the Federal Reserve money supply announcements, and to other observed changes in economic conditions.

The timing of a reaction to a shock is important when we wish to consider a policy that will create its own innovation; for example, how quickly and when will new plant and equipment expenditures occur is now, not only an important issue, but a solvable one. This should be considered when using vector autoregressions to estimate the impulse response of an economic system.

The further work to be done in this area involves both statistical procedures and the analysis of economic implications stemming from the use of these new procedures. A priority issue is to discover the approximate distribution for the identifiability test and to evaluate the rate of convergence of the estimates as well as to explore asymptotic normality of the estimators. In this connection more effort is needed to evaluate the power of the test at the sample sizes that are traditional in economic time series.

A major concern, of course, is whether improvements in forecasts can be obtained by non-linear forecasting procedures when confronted with non-invertible models. Preliminary work indicates that while the forecasts themselves are unchanged, non-invertibility has implications for the estimates of the confidence region associated with a given forecast.

On the economic side, the discovery of non-invertible models should stimulate considerable research into the economic significance of innovations. No longer should they be treated as a "black box", or as an uninteresting residuum for conventional analysis. Indeed, if an economic variable is an MA(q) process, then understanding the innovations is fundamental to understanding the economic series itself; the innovations are in fact more interesting economically than the actual series, because the innovations are the driving force behind the observed series. Further, the innovations may prove to be somewhat easier to analyze than the original observed series.

The results of this first paper should be easily extendable to full ARIMA processes, a topic which is currently under consideration.

Finally, in many areas of economics, the stock market, foreign trade, unemployment policy, and so on, the timing and initial time path of responses to policy shocks is of prime importance in evaluating alternative fiscal and monetary policy recommendations. Identification of the correct MA model will be crucial for this type of analysis.

REFERENCES

Brockwell P. and Davis R., (1987), Time series: theory and methods, Springer Verlag, New York.

Engle R., (1982), "Autoregressive conditional heteroscedasticity with estimates of the variance of the United Kingdom inflation", Econometrica, July, pp. 998-1007.

Granger C. and Newbold P., (1977), Forecasting economic time series, Academic Press.

Judge G. et al, (1985), The theory and practice of econometrics, John Wiley & Sons.

Lii K. and Rosenblatt M., (1982), "Deconvolution and estimation of transfer function phase and coefficients for non-Gaussian linear processes" Annals of Statistics, vol. 10, pp. 1195-1208.

Matsuoka T. and Ulrych T., (1984), "Phase estimation using the bispectrum", IEEE Proceedings, October.

Priestley M. B., (1981), Spectral Analysis and Time Series, Academic Press, London.

Rao S. and Gabr M., (1984), An introduction to bispectral analysis and bilinear models, Lecture Notes in Statistics, #24, Springer Verlag.