INTEGRATFD AUTOREGRESSIVE MOVING AVERAGE PROCESSES IN CONSUMPTION FUNCTIONS: A CRITIQUE ON ZELLNER'S et. al. PAPER

By THEODORE GAMALETSOS

Professor of Economics, Piraeus Graduate School of Industrial Studies

1. Introduction

The purpose of this paper is to critisize some hypotheses which have been used in the well known article by A. Zellner, D.S. Huang and L. C. Chau «Further Analysis of the Short-run Consumption Function with Emphasis on the Role of Liquid Assets» [8], and to apply an integrated autoregressive moving average process in these consumption functions using the same data.

2. Zellner's et. al. Paper: A review

In this paper Zellner et. al. report the results of additional experiments with the short - run consumption function. In particular they take up the problem of isolating expectation, inertia, and habit persistence effects and then they examine the problem of interpreting and estimating a real balance effect. In this work they estimated nonlinear distributed lag relationships by using nonlinear techniques. Further they examine the problem of autocorrelation in distributed lag schemes in a manner suggested by Fuller and Martin [4].

Aside from illustrating approaches to these methodological problems which they yielded results on the role of liquid assets in determining consumption expenditures which, they beliere are of consequence with respect to establishing direct influence of monetary variables on an expenditure relationship.

As regards expectation, inertia and habit persistence effects they consider the following range of hypotheses :

(1) $C_t = k_1 Y^e_t + u_{1t}$,

(2) $C_t = k_2 (Y^e_t + \beta Y^e_{t-1} + \beta^2 Y^e_{t-2} + \ldots + \beta^n Y^e_{t-n} + \ldots) + u_{2t}$,

and

(3) $C_t = \pi C_{t-1} + k_3 Y^e_t + u_{3t}$.

where C_t = quarterly real consumption, Y^e_t = quarterly real expected disposable income. The parameter β is introduced to represent possible inertia in reactions to changes in expected income.

Using Friedman's equation

(4) $Y^e_t - Y^e_{t-1} = (1 - \lambda) (Y_t - Y^e_{t-1})$,

where Y_t is price-deflated seasonally adjusted quanterly personal disposable income, and λ is a coefficient of expectations, equations (1), (2) and (3) become

(1,4) $C_t = \lambda C_{t-1} + k_t (1 - \lambda) Y_t + u_1 - \lambda u_{1t-1}$,

(2,4) $C_t = (\lambda + \beta) C_{t-1} - \lambda\beta C_{t-2} + k_2 (1 - \lambda) Y_t + u_{2t} - (\lambda + \beta) u_{2t-1} + \lambda\beta u_{2t-2}$

and

(3,4) $C_t = (\lambda + \pi) C_{t-1} - \lambda\pi C_{t-2} + k_3 (1 - \lambda) Y_t + u_{3t} - \lambda u_{3t-1}$.

Using U. S. quarterly data on C_t and Y_t, 1947 IV — 1961 II from Griliches et. al. [6] they concluded that neither (2,4) nor (3,4) is operative.

Their next step was to consider the role of liquid assets in affecting personal consumer expenditures. As a first approximation they formulate the consumption function as follows :

(5) $C_t = k_5 Y^e_t + \alpha (L_{t-1} - L^d_t) + u_{5t}$,

where L_{t-1} represents actual holding of real liquid assets at the beginning of the quarter, L^d_t represents the desired level of real liquid assets for the t[th] quarter and α is an adgustment coefficient ($\alpha > o$). About L^d_t they assume initially

(6) $L^d_t = \eta Y^e_t$,

a hypothesis in accord with Friedman's view of one of the main determinants of the demand for money. Combining (5) and (6) they obtain

$$(5,6) \qquad C_t = (k_5 - \alpha\eta)\, Y^e_t + \alpha L_{t-1} + u_{5t}\,.$$

Then by utilizing (4) and Koyck's transformation they obtained

$$(4,5,6) \qquad C_t = \lambda\, C_{t-1} + \alpha L_{t-1} - \alpha\lambda L_{t-2} + (k_5 - \alpha\eta)(1-\lambda)\, Y_t + u_{5t} - \lambda\, u_{5t-1}.$$

In this equation we observe that we have a problem of «overidentification». In this case we can estimate (4,5,6) subject to a certain constraint in these parameters.

As an alternative to (4) which may not incorporate trend considerations adequately, they consider

$$(7) \qquad Y^e_t - Y^e_{t-1} = (1-\lambda)\,(Y^e_t - Y^e_{t-1}) + \gamma\, Y^e_{t-1}\,,$$

where γ is the proportionate rate of increase of expected income ; that is, if $Y_t = Y^e_{t-1}$ then $Y^e_t - Y^e_{t-1} = \gamma\, Y^e_{t-1}$. On combining (7) with (5) and (6) they obtain

$$(5,6,7) \qquad C_t = (\lambda + \gamma)\, C_{t-1} + \alpha L_{t-1} - \alpha(\lambda+\gamma)\, L_{t-2} + (k_5 - \alpha\eta)(1-\lambda)\, Y_t + u_{5t} - (\lambda + \gamma)\, u_{5t-1}\,.$$

Since the interest rate may be a variable influencing desired liquid assets, they reformulate (6) to real :

$$(8) \qquad L^d_t = \eta\, Y^e_t - \delta\, i_t\,.$$

On compining (4), (5) and (8) they are led to the following results :

$$(4,5,8) \qquad C_t = \lambda\, C_{t-1} + \alpha\,(L_{t-1} - \lambda L_{t-2}) + \alpha\delta\,(i_t - \lambda\, i_{t-1}) + (k_5 - \alpha\eta)\,(1-\lambda)\, Y_t + u_{5t} - \lambda u_{5t-1}\,.$$

To test further the sensitivity of their estimates to specifying assumptions they combined (5), (7) and (8) to yield :

$$(5,7,8) \qquad C_t = (\lambda + \gamma)\, C_{t-1} + \alpha\,[L_{t-1} - (\lambda+\gamma)\, L_{t-2}] + \alpha\delta\,[i_t - (\lambda+\gamma)\, i_{t-1}] + (k_5 - \alpha\eta)\,(1-\lambda)\, Y_t + u_{5t} - \lambda\, u_{5t-1}\,.$$

In terms of their equation (4,2,8) they introduced the following hypothesis ;

$$(9) \qquad v_t = \rho\, v_t + \varepsilon_t\,,$$

where $v_t = u_{5t} - \lambda u_{5t-1}$ and ε_t is assumed to be a «classical» disturbance. On combining (9) and (4,5,8) they obtain

$$(4,5,8,9) \qquad C_t = (\lambda+\rho)\, C_{t-1} - \rho\lambda\, C_{t-2} + \alpha\,(L_{t-1} - \lambda L_{t-2}) - \rho\alpha\,(L_{t-2} - \lambda L_{t-3} + \\ + \alpha\delta\,(i_t - \lambda i_{t-1}) - \rho\alpha\delta\,(i_{t-1} - \lambda i_{t-2}) + (k_5 - \alpha\eta)(1-\lambda)(Y_t - \rho Y_{t-1}) + \\ + \varepsilon_t\,.$$

From their estimates they found that U.S. quarterly data 1947 IV—1961 II do not support that inertia or habit persistence effects are operative **in addition** to an income expectation effect. They conclude also that the hypothesis that imbalances in consumer liquid asset holdings exert a statistically and economically significant influence on consumption. This is important since it constitutes an evidence that monetary variables affect an important expenditure relationship directly and not just indirectly through interest rate effects.

3. Comments on Zellner's et. al. Paper

α) If we consider instead of (3) a pure habit persistence hypothesis, say

$$(3^*) \qquad C_t = \pi\, C_{t-1} + k_3 Y_t + u_{3t}$$

then we will have as a result that (3*) will be indistinguishable from (1,4), which means that the hypothesis (3*) alone can explain the behavior of consumer expenditures without the need to use together the expectation hypothesis, and vise versa.

β) Zellner et. al. suggested that neither (2,4) nor (3,4) is operative. They reached to this conclusion by testing the coefficient of C_{t-2} variable, which according to the data used is found to be not significantly different from zero.

The effect however of ignoring possible autocorrelation in the disturbances in equations (2,4) and (3,4) could lead to a biased estimation of the coefficient of C_{t-2}, which bias could lead to doubt about the rejection of the hypothesis that this coefficient is statistically insignificant [1].

γ) The estimated $\hat{\lambda}$ in the equation (4,5,8) is about the same as the estimated $(\hat{\lambda} + \hat{\rho})$ in the equation (4,5,8,9). This means that $\hat{\rho} \simeq 0$, which implies that the assumption (9) is not the correct one.

δ) The assumption (9) that $(1 - \rho L)\, v_t = \varepsilon_t$ is random implies that

1) As we know, we obtain a bias estimation of the parameters in an equation where we assume autocorrelation among the disturbances and within the regressors we have the regressant with some lag period. For this see T. Gamaletsos [5] pp. 239 - 247.

$(1 - \rho L)\ (1 - \lambda L)\ u_t = \varepsilon_t$ is random, that is we have a second - order autoregressive process for u_t in equation (5), which is of the fom

$$u_{5t} = (\rho + \lambda)\ (Lu_t) - \rho\lambda\ (L^2 u_t) + \varepsilon_t ,$$

where L is the lag operator.

The non-linear two-stage leart squares estimates of λ and ρ are respectively $\hat{\lambda} \simeq 0{,}8$ and $\hat{\rho} \simeq -0{,}3$, which gives approximately

$$u_{5t} = {}^1/_2\ (L\ u_t) + {}^1/_4\ (L^2 u_t) + \varepsilon_t .$$

ε) The non-linear classical least squares estimation and the non-linear two-stages least squares estimation methods affect the pattern of distributed lags in the model (4,5,8,9) [2].

στ) The equations (1,4), (2,4) and (3,4) which have been estimated, are derived by applying hypothesis (4) on the equations (1), (2) and (3) respectively.

The equation (4) — or the equation (7) — is derived by assuming that the discrete distribution of τ (the lag patterns) is of the form $w_\tau = (1 - \lambda)\ \lambda^\tau$, where $0 < \lambda < 1$. However this form of distribution of lags, called «geometric distribution», proposed by Koyck, is one among many others and there is no **a priori** necessity to use that special distribution. We should first test which is the appropriate distribution to apply in this case. But before we do this a brief review of the theory of distributed lag models is needed.

4. The Theory of Distributed Lag Models : A Review

Generally in the equations of the form

$$\text{(10)} \qquad y_t = \beta\ (w_0\ x_t + w_1\ x_{t-1} + w_2\ x_{t-2} + \ldots) + u_t$$

or

$$\text{(11)} \qquad y_t = \beta w\ (L)\ x_t + u_t$$

by using the assumptions $w_\tau \geqslant 0$ and $\Sigma_{\tau=0}^{\infty}\ w_\tau = 1$ we can note that w_τ ' s have the properties of a discrete probability distribution defined over the integers $\tau = 0, 1, 2, \ldots$; this leads to the need of discussing the form of the distributed lags τ and finding the moments of these distributions — in the case we are not interested for the whole distribution. The sequence of w_τ ' s describes the **form** of the lag, the «time path of an economic reaction».

---

2) For patterns of distributed lags see Griliches [7].

There are some advantages in the use of the notion of probability distribution. In equation (10) or (11) above we can interpreat w(L) as a polynomial in the lag operator and as the **lag-genereting function** [3]. This interpretation has the advantages of a probability generating function from which we can get the moments of the distribution w_τ. For example by using the first and second derivatives of this function at a specific point we derive easily the first and second moments of that distribution.

Many patterms of distributed lags are possible ; any probability distribution over the nonnegative integers is available. And we need to have **a priori** knowledge of the distribution of w_τ for estimation and for analytical purposes.

Below we will present briefly the most important distributions of w_τ, which have been used by several researchers in applied economics.

α) Finite Distribution

In this case we have $w_\tau = 0$ for $\tau > h$, where h is an integer. The model (10) in that case, omitting the disturbance, can be written as $y_t = \beta\, w(L)\, x_t$, where $w(L) = \Sigma_{\tau=0}^{\infty} w_\tau$.

β) Log-normal Distribution (proposed by Fisher)

According to this distribution w_τ is the probability that a normal variable with mean μ and variance σ^2 takes a value between $\ln\tau$ and $\ln(\tau + 1)$.

γ) The Distribution $w_\tau = \lambda^2 \tau e^{-\lambda\tau}$ (proposed by Theil and Stern).

δ) The Geometric Distribution (proposed by Koyck)

According to this distribution we have $w_\tau = (1 - \lambda)\, \lambda_\tau$ where $0 < \lambda < 1$. The model (10) in this case, omitting the disturbance, takes the form

$$y_t = \beta\,(1 - \lambda)\,(1 - \lambda L)^{-1}\, x_t .$$

The mean of this one-parameter distribution is $E(\tau) = \lambda / (1 - \lambda)$ and the variance is $v(\tau) = \lambda / (1 - \lambda)^2$.

ε) Pascal or Negative Binomial Distribution (proposed by Solow)

---

3) If the function $A(z) = \alpha_0 + \alpha_1 z + \alpha_2 z^2 + a_3 z^3 + \ldots$ — where z is a dummy variable — has a limit, then A(z) is called the generating function of the sequence $\{\alpha_i\}$. In addition if all $\alpha_i \geqslant 0$ and $A(1) = 1$, i.e $\Sigma_{i=1}^{\infty} \alpha_i = 1$, then A(z) is a probability generating function.

In this distribution we have $w_\tau = \binom{r+\tau-1}{\tau}(1-\lambda)^r \lambda^\tau$, where $0 < \lambda < 1$ and r is a positive integer. Assuming this distribution the model (10) takes the form

$$y_t = \beta\,(1-\lambda)^r\,(1-\lambda L)^{-r}\,x_t + u_t\,.$$

This two parameter distribution permits wide variety of shapes ; for $r = 1$ it becomes a geometric distribution.

στ) **Rational Distribution** (proposed by Jorgenson)

When the w(L) can be factorised into a ratio of two finite polynomials, say w(L) = U(L) / V(L), then we have a rational distribution. In this case the model (10) can be written as $V(L)\, y_t = U(L)\, x_t$ (omitting the disturbance). We observe that the rational distribution is a more general case than the negative binomial distribution, which in turn is a more general case than the geometric distribution.

It is worthwile to notice that Jorgenson's proposal is very similar with that of Box and Jenkins [4].

Griliches [7] suggests that the difference equation $V(L)\, y_t = U(L)\, x_t$ has a bounded solution for arbitraty initial conditions if and only if V(L) is stable, that is if the associated characteristic equation $z^n T(z^{-1}) = 0$ has all its toots inside the unit cycle — which is a similar idea with that of Box and Jenkins [5]. And for the resulting sequence to be an acceptable distributed lag function (i.e. nonnegative and convergent) it is sufficient for **both** sequences V(L) and U(L) to be convergent and nonnegative. If U(L) is a constant — as it happens in the models (2,4), (3,4) and (4,5,8,9) — we then examine only V(L).

A necessary condition for $V(L)^{-1}$ to be convergent and nonnegative is that the maximal root of the associated characteristic equation $V(Z^{-1}) = 0$ be possitive and les than one. If we want a well behaved «smooth» lag distribution, it is sufficient that the roots of $V(Z^{-1}) = 0$ are real and positive [6].

---

4) Compare the above equation $V(L)y_t = U(L)x_t$ with $\Phi(B)z_t = \Theta(B)\alpha_t$ of Box and Jenkins [2] p. 11.

5) Box and Jenkins [2] define an autocovariance generating function $C(B) = \sum_{k=-\infty}^{+\infty} \gamma_k B^k$ and this function converges for $|B| < 1$, and since the autocovariance generating of a linear process factorizes $C(B) = L(B)\, L(B^{-1})$, the above condition implies that L(B) must converge for $|B| \,|<|\, 1$, that is, within the unit cycle.

6) The constraints on the admissible range of the coefficients of V(L) are sufficient but not necessary.

If the roots of $V(Z)^{-1} = 0$ are real, positive and distinct, then $V(L)^{-1}$ can be written as a convolution of a number of geometrically declining lag distributions, the number convoluted being equal to the number of roots of $V(Z^{-1}) = 0$.

In the (general) Pascal distribution the roots of $V(Z^{-1}) = 0$ will be real positive and equal. But in a Rational distributed lag function we have not this constraint of Pascal distribution ; i.e. equality of the roots. However the resulting difference equation may still imply an acceptable distributed lag function.

As we mentioned before, when we know the form of the distribution of w_τ we can estimate its moments. For examble, when $w_\tau = (1 - \lambda)\lambda^\tau$ then the mean is $E(\tau) = \lambda / (1 - \lambda)$ and the variance is $V(\tau) = \lambda / (1 - \lambda)^2$;when $w(L) = U(L) / V(L)$, where $U(L) =$ constant, and $V(L) = (1 - bL - cL^2)$ the mean is $E(\tau) = (b + 2c) / (1 - b - c)$. Griliches says that this $E(\tau)$ may not be so sensitive to slight changes in b and c. Uncertainty about b and c separately can imply uncertainty about the shape of the lag distribution, but not necessarily about the average lag $E(\tau)$.

The average lags of the models (2,4) and (3,4) are respectively

$$E_1(\tau) = \frac{\pi}{1-\pi} + \frac{\lambda}{1-\lambda} \text{ and } E_2(\tau) = \frac{\beta}{1-\beta} + \frac{\lambda}{1-\lambda},$$

while the average lag of the model (4,5,8,9) is

$$E_3(\tau) = \frac{\rho}{1-\rho} + \frac{\lambda}{1-\lambda}.$$

Following Griliches, to have a nonnegative lag distribution in the model (4.5,8,9) the parameters b and c must satisfy the following restrictions : α) $0 < b < 2$, β) $-1 < c < 1$, γ) $(1 - b - c > 0$, and δ) $b^2 \geqslant -4c$, where $b = (\lambda + \rho)$ and $c = -\lambda\rho$. For $b^2 = -4c$ we have Pascal distribution. Zellner et. al. using non-linear classical least squares and non-linear two-stages least squares estimation methods have found $\hat{\rho}_c \simeq -0{,}10$, $\hat{\lambda}_c \simeq 0{,}33$ and $\hat{\rho}_T \simeq -0{,}29$, $\hat{\lambda}_T \simeq 0{,}82$, where $\hat{\rho}_c$ and $\hat{\lambda}_c$ are the estimates of ρ and λ using NL—CLS method, while $\hat{\rho}_T$ and $\hat{\lambda}_T$ are the estimates of ρ and λ using NL — TSLS estimation method. Therefore using NL — CLS method we optain $\hat{b}_c = (\hat{\lambda}_c + \hat{\rho}_c) \simeq 0{,}23$ and $\hat{c}_c = - \hat{\rho}_c \hat{\lambda}_c = - (0{,}10)\ (0{,}33) = 0{,}03$, while using NL — TSLS method for the same parameters we obtain the estimates $\hat{b}_T \simeq 0{,}82 - 0{,}29 = 0{,}53$ and $\hat{c}_T = - (0{,}29)\ (0{,}82) = 0{,}24$.

Therefore these estimates of b and c using NL — CLS and NL — TSLS estimation methods satisfy the above restrictions, i.e. we have a nonnegative lag distribution in the model (4,5,8,9).

The average lag accosiated with the NL — CLS estimates is

$$E(\hat{\tau}) = \frac{\hat{\rho}_c}{1-\hat{\rho}_c} + \frac{\hat{\lambda}_c}{1-\hat{\lambda}_c} = \frac{0{,}29}{0{,}74} \simeq 0{,}4 \text{ quarters},$$

while this average lag accosiated with the NL—TSLS estimates is

$$E(\hat{\tau}) = \frac{\hat{\rho}_T}{1-\hat{\rho}_T} + \frac{\hat{\lambda}_T}{1-\hat{\lambda}_T} = \frac{1{,}01}{0{,}23} = 4{,}4 \text{ quarters.}$$

Therefore we observe that the average lag of the model (4,5,8,9) changes very much by using NL – CLS and NL — TSLS estimation procedures.

Jorgenson's Rational distribution is analogous to Box—Jenkins's proposal for an ARMA (p,q) — a q – order moning average and p — order autoregressive scheme — which is adequate to describe most **stationary** time series, for small p and q.

However the ARMA (p, q) process is a special case of the IARMA (p, d, q) the integrated ARMA (p, d, q) process, which, for appropriate choice of the parameters, fits observed **non - stationary** time series [7].

Now Box and Jenkins ARMA (p, q) model is given by the formula

$$(12) \qquad \varphi(B)\,\dot{z}_t = \theta(B)\,\alpha_t\,,$$

where φ (B) and θ (B) are polynomials in B of degree p and q respectively. To ensure stationarity the roots of the characteristic equation φ (B) = 0 must lie inside the unit cycle [8]. Therefore a natural way of obtaining non - stationary processes is to relax this restriction. For non-stationary time series, it appears that the ARMA (p, q) process will fit the first or second or d-th order difference of the $\dot{z}$ series. This is the IARMA (p, d, q) and the model (12) becomes

7) Note that ARMA (d, q) = IARMA (0, d, q).

8) For example in the model $(1 - \varphi B)\dot{z}_t = \alpha_t$, which is an ARMA (1, 0), to ensure stationarity we must have $|\varphi| < 1$. When $|\varphi| > 1$ we have non - stationarity.

$$(13) \qquad \varphi(B)\, \Delta^d z_t = \theta(B)\, \alpha_t .$$

All these models are linear, but this should not bother us since we can transform some nonlinear models into a linear one by a suitable transformation, say a logarithmic transformation. The possibility of improvement of model adequacy by transformation should always be borne in our mind.

Now the question that arises is which of the previous models best describes a given time series. If we know the pattern of the distributed lags we can use it for forecasting ; and what we are interested in here is to forecast and not to estimate the structural parameters of the above described models.

The «permanent», which we better call «expected» income hypothesis (1) in Zellner et. al. assumes that a consumer is attempting to forecast his income on the basis of the past values of the observed income. The measured income is a time series, a stochastic process, the form of which we are trying to find out. When we know the stochastic (stationary or nonstationary) process which generates measured income, say Y_t, we can find its optimal predictor (forecast), say Y_t^e [9].

Friedman's optimal predictor (4) or

$$(14) \qquad Y_t^e = \beta \sum_{\tau=1}^{\infty} (1-\beta)^{\tau-1}\, Y_{t-\tau} = \beta \sum_{\tau=0}^{\infty} (1-\beta)^{\tau}\, Y_{(t-1)-\tau}$$

is an «exponentially» weighted average of past observations [10].

Box and Jenkins give a more general form of Friedman's predictor, which is of the form

$$(15) \qquad \Delta\, Y_{t+1}^e = \left[\sum_{i=1}^{i=1} \gamma_{-i}\, S^{-i} + \sum_{k=0}^{m} \gamma_k\, S^k \right] e_t$$

9) By an optimal predictor we mean the minimum variance unbiased predictor, which is given by the conditional expectation $E(Y_t \mid Y_{t-1}, Y_{t-2}, \dots)$.

10) Equation (14) is obtained from (4) by using the lag operator as follows :

$$[\,1-(1-\beta)\, L\,]\; Y_t^e = \beta\, Y_{t-1}$$

or

$$Y_t^e = [\,1-(1-\beta)\, L\,]^{-1}\, \beta Y_{t-1}$$

or

$$Y_t^e = [\,1 + (1-\beta)\, L + (1-\beta)^2 L^2 + \dots\,]\, \beta\, Y_{t-1}$$

$$= \beta Y_{t-1} + \beta(1-\beta)\, Y_{t-2} + \beta(1-\beta)^2 Y_{t-3} + \dots$$

$$= \beta \sum_{\tau=1}^{\infty} (1-\beta)^{\tau-1}\, Y_{t-\tau} .$$

where $S^{-i} e_t = \Delta^j e_t$ (the j th difference of e_t) and $e_t = Y_t - Y^e_t$; this is an «adaptive» forecast, because it depends on the past forecast.

The predictor (25) above is optimal for a stochastic process, which is generated from

$$(16) \qquad \Delta Y_{t+1} = \Delta \delta_{t+1} + \Delta Y^e_{t+1}$$

$$= \Delta \delta_{t+1} + [\Sigma_{i=-1}^{i=1} S^{-i} + \Sigma_{k=0}^{m} \gamma_k S^k] e_t$$

$$= \Delta \delta_{t+1} + \Sigma_{j=-1}^{m} \gamma_j S^j \delta_t$$

where δ_t's are non-autocorrelated identically distributed random variables with $E(\delta_t) = 0$, i.e. they are «random deviates».

If the stochastic process (16) is non-stationary we can transform it into stationary by differencing m times, if the population serial covariances of lag greater than $m + l + 1$ were zero ; and the predictor (15) would be an optimal one.

Differencing (16) m times we obtain

$$(17) \qquad \Delta^{m+1} Y_{t+1} = \Delta^{m+1} \delta_{t+1} + \Sigma_{j=0}^{l+m} \gamma_j \Delta^{l+m-j} \delta_t$$

which we rewrite as

$$(18) \qquad \Delta^{m+1} Y_{t+1} = \delta_{t+1} + \Sigma_{j=0}^{l+m} \eta_j \delta_{t-j}$$

so that all serial covariances of lag $l + m + 1$ and after are zero. Equation (18) now represents a stationary moning average process of order $l + m + 1$.

Note that Friedman's predictor corresponds simply to the central term in the general series (15), namely $\Delta Y^e_{t+1} = \gamma_0 e_t$ (or $\Delta Y^e_t = \gamma_0 e_{t-1}$), which can be written in a more familiar formula as

$$(19) \qquad Y^e_{t+1} = \gamma_0 \Sigma_{j=0}^{\infty} (1 - \gamma_0)^j Y_{t-j}$$

or for $\gamma_0 = 1 - \lambda$ we have

$$(20) \qquad Y^e_{t+1} = (1 - \lambda) \Sigma_{j=0}^{\infty} \lambda^j Y_{t-j}$$

$$= (1 - \lambda) (1 - \lambda L)^{-1} Y_t \qquad ,$$

which is hypothesis (4) in Zellner et. al.

Therefore the general form (15) of Friedman's predictor comes by adding «past» and «future» terms in the well-known exponential form (20). This equation (20) is an optimal predictor for the stochastic process

$$(21) \qquad Y_{t+1} = m + \delta_{t+1} + \gamma_0 S^1 \delta_t ,$$

for which the first difference is a first - order moning average process.

In the general form (16) if $\gamma_i = 0$ for all $i \neq 0$, then this equation becomes

$$(22) \qquad \Delta Y_{t+1} = \gamma_0 \delta_t + \Delta \delta_{t+1} .$$

Thus if (22) is correct then $\rho_\tau = 0$ for all $\tau \geqslant 2$, where ρ_τ is the coefficient of serial correlation of lag τ. This provides a convenient check up on model (22). The optimal predictor for this model, as we mentioned above, is equation (20). If $\gamma_i = 0$ for all $i \neq 0$ and $i \neq -1$ equation (16) becomes

$$(23) \qquad \Delta Y_{t+1} = \gamma_0 \delta_t + \Delta \delta_{t+1} + \gamma_{-1} \Delta \delta_t ,$$

which it is correct then $\rho_\tau = 0$ for all $\tau \geqslant 3$. The associated optimal predictor is given by the form

$$(24) \qquad \Delta Y^e_{t+1} = \gamma_{-1} \Delta e_t + \gamma_0 e_t$$

which is a special case of (15). If $\gamma_i = 0$ for all $i \neq 0, 1, -1$, then (16) becomes

$$(25) \qquad \Delta Y_{t+1} = \gamma_0 \delta_t + \Delta \delta_{t+1} + \gamma_{-1} \Delta \delta_t + \gamma_1 \Sigma_{j=0}^{\infty} \delta_{t-j} .$$

Now if (25) is correct then $\rho_\tau = 0$ for all $\tau \geqslant 4$. The corresponding optimal predictor in this case is

$$(26) \qquad \Delta Y^e_{t+1} = \gamma_0 e_t + \gamma_{-1} \Delta e_t + \gamma_1 \Sigma_{j=0}^{\infty} e_{t-j} .$$

The second difference of (25) is a moving average process of order 3, that is

$$(27) \quad \Delta^2 Y_{t+1} = \delta_{t+1} + (\gamma_{-1} + \gamma_0 + \gamma_{-1} - 2)\, \delta_t + (1 - 2\gamma_{-1} - \gamma_0)\, \delta_{t-1} + \gamma_{-1} \delta_{t-2} .$$

Note that the change in the predictor in any period t is a function of there separate «controls» : Δe_t, e_t and $\Sigma_{j=0}^{\infty} e_{t-j}$; Box and Jenkins named them «first difference» term or «derivative control», «proportional» term and «cumulative» term or «integral control» respectively.

The selection of the appropriate model is of course vital. And what we have to do here is to identify the model, which best describes the given time series (which is the same with that used by Zellner et. al., i.e. disposable income in billions of 1954 dollars, seasonally adjusted at annual rates, 1947 IV—1961 II, taken by Griliches et. al.).

5. Model Identification

What we need first of all in this case is to know how many times we will difference the non-stationary time series to produce stationarity. After that we have to determine the degree of the polynomials $\varphi(B)$ and $\theta(B)$ in equation (13) above, i.e. how many terms will be included in the IARMA (p, d, q) model fitting this stationary series.

5.1. Use of the Autocorrelation Function

The test procedure here for specifying the appropriate IARMA (b, d, q) model is to compare the estimated autocorrelation coefficient $\hat{\rho}_\tau$ with its standard error under the assumption that the process is a moving average of order $(\tau - 1)$. For values of $\tau > q$ — where q is the degree of the polynomial $\theta(B)$ — $\hat{\rho}_\tau$ should be small compared with its standard error.

Using Bartlett's formula

$$(28) \qquad V(\rho_\tau) \simeq \frac{1}{T}\left[1 + 2\,\Sigma_{k=1}^{t-1}\,\rho_k^2\right]$$

and replacing ρ_τ by $\hat{\rho}_\tau$ we have

$$(29) \qquad V(\hat{\rho}_\tau) \simeq \frac{1}{T}\left[1 + 2\,\Sigma_{k=1}^{\tau-1}\,\hat{\rho}_k^2\right].$$

Now if the process is of order τ, the statistic

$$(30) \qquad s_\tau = \frac{\hat{\rho}}{\sqrt{V(\hat{\rho}_\tau)}}$$

will approximately be distributed as a unit normal variable. And if a series of s values beyond some point q lie between the 95 % limits ± 1,96, it can be concluded that the process is a MA (q).

Table 1 shows the first twelve estimated autocorrelation coefficients of

TABLE 1

τ	$\hat{\rho}_\tau$	$\hat{\rho}^2_\tau$	$\Sigma_{k=1}^{\tau-1} \hat{\rho}^2_\tau$	$\left(1 + 2\,\Sigma_{k=1}^{\tau-1} \hat{\rho}^2_\tau\right)$	$V(\hat{\rho}_\tau)$	$\sqrt{V(\hat{\rho}_\tau)}$	$\lvert s_\tau \rvert$
1	—0,03	0,0009	—	1	0,01820	0,1349	0,2224
2	—0,05	0,0025	0,0009	1,0018	0,01823	0,1350	0,3704
3	—0,06	0,0036	0,0034	1,0068	0,01832	0,1354	0,4431
4	—0,24	0,0576	0,0070	1,0140	0,01845	0,1358	1,7673
5	—0,14	0,0196	0,0646	1,1292	0,02055	0,1434	0,9763
6	—0,11	0,0121	0,0842	1,1684	0,02126	0,1458	0,7545
7	0,07	0,0049	0,0963	1,1926	0,02171	0,1474	0,4752
8	—0,10	0,0100	0.1012	1,2024	0,02188	0,1479	0,6761
9	—0.02	0,0004	0,1112	1,2224	0,02220	0,1490	0,1342
10	—0,02	0,0004	0,1116	1,2232	0,02230	0,1493	0,1340
11	0	0	0,1120	1,2240	0,02230	0,1493	0
12	0,30	0,0900	0,1120	1,2240	0,02230	0,1493	2,009*

the first differences of Y_t with their estimated standart deviations and their statistic s_τ. From these estimates we see that all $|s_\tau|$ values, with only one exception, lie within the 95% limits. It could therefore be correctly concluded that the process $\{Y_t\}$ is of zero-order, i.e. that the appropriate model is the $\varphi(B) z_t = \theta(B) \alpha_t$, where $\varphi(B)$ is a zero degree polynomial.

5.2. Use of the Partial Autocorrelation Function

Box and Jenkins suggest that if the autocorrelation function and partial autocorrelation function exhibit opposite behavior we have a pure autoregressive or a pure moving average process; but if both autocorrelation functions tail off we conclude that we have an ARMA (p, q) process.

To estimate the partial autocorrelation coefficients we use the recursive formulae, which is due to Durbin :

$$(31) \qquad \hat{\varphi}_{\tau+1,j} = \hat{\varphi}_{\tau j} - \hat{\varphi}_{\tau+1,\tau+1} \cdot \hat{\varphi}_{\tau,\tau-j+1} \qquad (j = 1, 2, \ldots, \tau),$$

and where

$$(32) \qquad \hat{\varphi}_{\tau+1,j} = \frac{\hat{\rho}_{\tau+1} - \sum_{j=1}^{\tau} \hat{\varphi}_{\tau j} \hat{\rho}_{\tau+1-j}}{1 + \sum_{j=1}^{\tau} \hat{\varphi}_{\tau j} \hat{\rho}_j}\,.$$ [11]

We have an A R (k — 1) process if a sequence of u'_k s beyond the k — 1 lie within the 95% limits ± 1,96, where

$$(33) \qquad u_k = \frac{\hat{\varphi}_{kk}}{\sqrt{V(\hat{\varphi}_{kk})}}$$

is distributed normally and where

$$(34) \qquad V(\hat{\varphi}_{kk}) \simeq \frac{1}{T-k}\,.$$

By using the estimated first difference $\hat{\rho}_\tau$'s we obtain the first $\hat{\varphi}_{kk}$'s (for k = 1, 2, 3, 4, 5), i.e.

---

11) Note that $\hat{\varphi}_{11} = \hat{\rho}_1$.

$$\hat{\varphi}_{11} = \hat{\rho}_1 = -0{,}03$$

$$\hat{\varphi}_{22} = \frac{\hat{\rho}_2 - \hat{\rho}^2_1}{1 - \hat{\rho}^2_1} = \frac{-0{,}05 - 0{,}0009}{1 - 0{,}009} = -0{,}05095 \text{ ,}$$

$$\hat{\varphi}_{33} = \frac{\hat{\rho}_3 - \hat{\varphi}_{21}\hat{\rho}_2 - \hat{\varphi}_{22}\hat{\rho}_1}{1 + \hat{\varphi}_{21}\hat{\rho}_1 + \hat{\varphi}_{22}\hat{\rho}_1} = -0{,}06289 \text{ ,}$$

where

$$\hat{\varphi}_{21} = \frac{\hat{\rho}_1(1 - \hat{\rho}_2)}{1 - \hat{\rho}_1^2} = -0{,}03153 \text{ ,}$$

$$\hat{\varphi}_{44} = \frac{\hat{\rho}_4 - \hat{\varphi}_{31}\hat{\rho}_3 - \hat{\varphi}_{32}\hat{\rho}_2 - \hat{\varphi}_{33}\hat{\rho}_1}{1 + \hat{\varphi}_{31}\hat{\rho}_1 + \hat{\varphi}_{32}\hat{\rho}_2 + \hat{\varphi}_{33}\hat{\rho}_3} = -0{,}24479 \text{ ,}$$

where

$$\hat{\varphi}_{31} = \hat{\varphi}_{21} - \hat{\varphi}_{33}\hat{\varphi}_{22} = -0{,}03473$$

and

$$\hat{\varphi}_{32} = \hat{\varphi}_{22} - \hat{\varphi}_{33}\hat{\varphi}_{21} = -0{,}05293 \text{ ,}$$

and finally

$$\hat{\varphi}_{55} = \frac{\hat{\rho}_5 - \hat{\varphi}_{41}\hat{\rho}_4 - \hat{\varphi}_{42}\hat{\rho}_3 - \hat{\varphi}_{43}\hat{\rho}_2 - \hat{\varphi}_{44}\hat{\rho}_1}{1 + \hat{\varphi}_{41}\hat{\rho}_1 + \hat{\varphi}_{42}\hat{\rho}_2 + \hat{\varphi}_{43}\hat{\rho}_3 + \hat{\varphi}_{44}\hat{\rho}_4} = -0\,1563 \text{ ,}$$

where

$$\hat{\varphi}_{41} = \hat{\varphi}_{31} - \hat{\varphi}_{44}\hat{\varphi}_{33} = -0{,}05012 \text{ ,}$$

$$\hat{\varphi}_{42} = \hat{\varphi}_{32} - \hat{\varphi}_{44}\hat{\varphi}_{32} = -0{,}06589 \text{ ,}$$

$$\hat{\varphi}_{43} = \hat{\varphi}_{33} - \hat{\varphi}_{44}\hat{\varphi}_{31} = -0{,}7139 \text{ .}$$

In the the following Table 2 we report the estimates u_k for $k = 1,2,3,4,5$.

TABLE 2

τ	$\hat{\varphi}_{kk}$	$\sqrt{V(\hat{\varphi}_{kk})}$	u_k
1	— 0,03000	0,13609	— 0,2204
2	— 0,05095	0,13736	— 0,3709
3	— 0,06289	0,13868	— 0,4535
4	— 0,24479	0,14003	— 1,7481
5	— 0,15630	0,14142	— 1,1052

From this table we observe that all u_k ' s lie within the 95% limits and we conclude that an AR (0) process is adequate to fit the given time series. The θ (B) is a zero degree polynomial, that is all θ's are zero, and this is equivalent to say that in the model (32) $\gamma_0 = 1$.

Comparing the results of sections 4.1 and 4.2 above we conclude that an IARMA (0, 1, 0) process fits our data. Therefore the model $\varphi(B)\, z_t = \theta(B)\, \alpha_t$ in this case becomes $(1-L)\, Y_t = \alpha_t$ or $\Delta Y_{t+1} = \alpha_t$, and the optimal predictor is $\Delta Y^e_{t+1} = \gamma_0 e_t = e_t$, i.e. $Y^e_{t+1} = Y_t$ or $Y^e_t = Y_{t-1}$.

This means that the expected value of the disposable income at period t equals with the value of the last period measured income [12].

Since the autocorrelation and partial autocorrelation functions have not a cut - off point, it means that our model is a mixed IARMA (p, d, q) process with $p = q = 0$ and $d = 1$.

4.3 Liquidation of the Autocorrelation Function

The autocorrelations ρ (α) of the α - process are related to the autocorrelations ρ_τ (Y) of the Y - process by the relationship

12) That this is the appropriate model it can easily be seen in the Diagram 2 (in the Appendix) where almost without exception all the first twenty four $\hat{\rho}_\tau$ ' s are between $\pm$ 0,20, which is an approximation of the standard deviation $\sqrt{V(\hat{\rho}_\tau)} \simeq 0,14$. The fluctuations of the $\hat{\rho}_\tau$ function show, according to Box and Jenkins's terminology, a «white noise».

$$(35)\qquad \rho_\tau(\alpha) = k^{-1} \frac{\varphi(B)\,\varphi(B^{-1})}{\theta(B)\,\theta(B^{-1})}\,\rho_\tau(Y)$$

where

$$k = \frac{\varphi(B)\,\varphi(B^{-1})}{\theta(B)\,\theta(B^{-1})}\,\rho_0(Y).$$

Now because in our model $\varphi(B) = \theta(B) = 1$ we have $k = \rho_0(Y)$ and $\rho_\tau(\alpha) = \rho_\tau(Y) / \rho_0(Y)$; Using the estimated autocorrelation coefficients $\hat{\rho}_\tau(Y)$ of the Y - process we easily find the estimated autocorrelation coefficients of the α - process (the residual process) by $\hat{\rho}_\tau(\alpha) = \hat{\rho}_\tau(Y) / \hat{\rho}_0(Y)$, which shows that the $\hat{\rho}_\tau(\alpha)$ function is very similar to $\hat{\rho}_\tau(Y)$ function. This means that the operator $\theta^{-1}(B)\ \varphi(B) = 1$ converts our Y - process to a white noise.

Another, rather empirical, way of testing that in the model $(1-L)Y = [1-\theta(B)]\alpha_t$ we have $\theta=0$ is to use different values, between 1 and -1, for the parameter θ in the equation $\hat{\alpha}_t = \hat{\theta}\hat{\alpha}_{t-1} + Y_t - Y_{t-1}$ and to take that value of $\hat{\theta}$ which minimizes $\Sigma\,\hat{\alpha}_t^2$. The minimum $\Sigma\,\hat{\alpha}_t^2$ will be given by using $\hat{\theta} = 0$ (or $\gamma_0 = 1$).

6. Conclusion

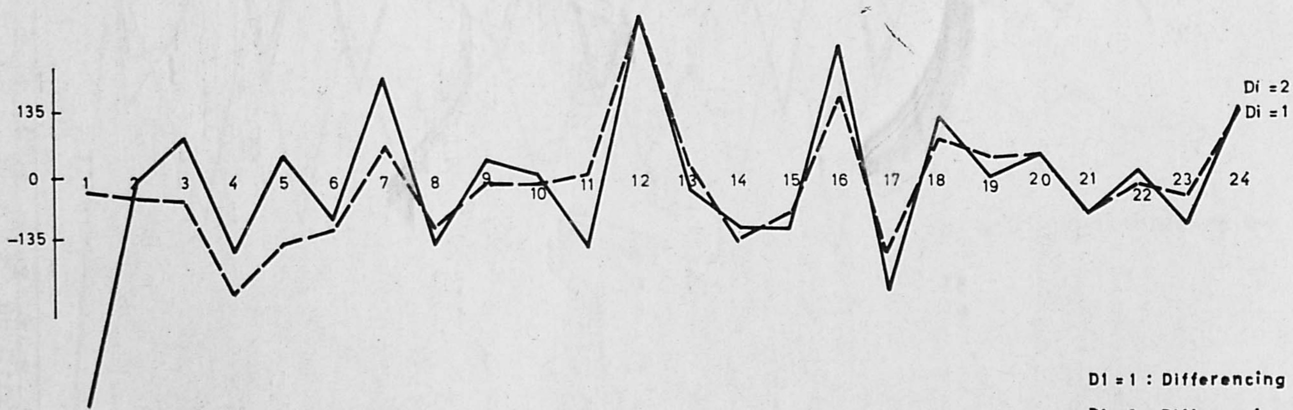
From the results of the previous section 4 we conclude that the model $\Delta Y_{t+1} = \alpha_t$ fits our data and therefore the optimal predictor is $Y^e_t = Y_{t-1}$.

Now if this is the case then equations (1), (2) and (3) in Zellner et. al. paper can be directly estimated. The advantage of the direct estimation is that we can estimate these equations by using the classical least squares estimation method ; we may have a loss of efficiency but not inconsistency when we assume autocorrelation of the direct disturbances, which is the more probable case in the time series.

APPENDIX

DIAGRAM 1

Estimated Autocorrelation Coefficients $\hat{\rho}_\tau$ (first and second difference) with their standard errors



D1 = 1 : Differencing once

D1 = 2 : Differencing twice

DIAGRAM 2

Estimated Autocorrelation Coefficients $\hat{\rho}_\tau$ (third and forth difference) with their standard errors

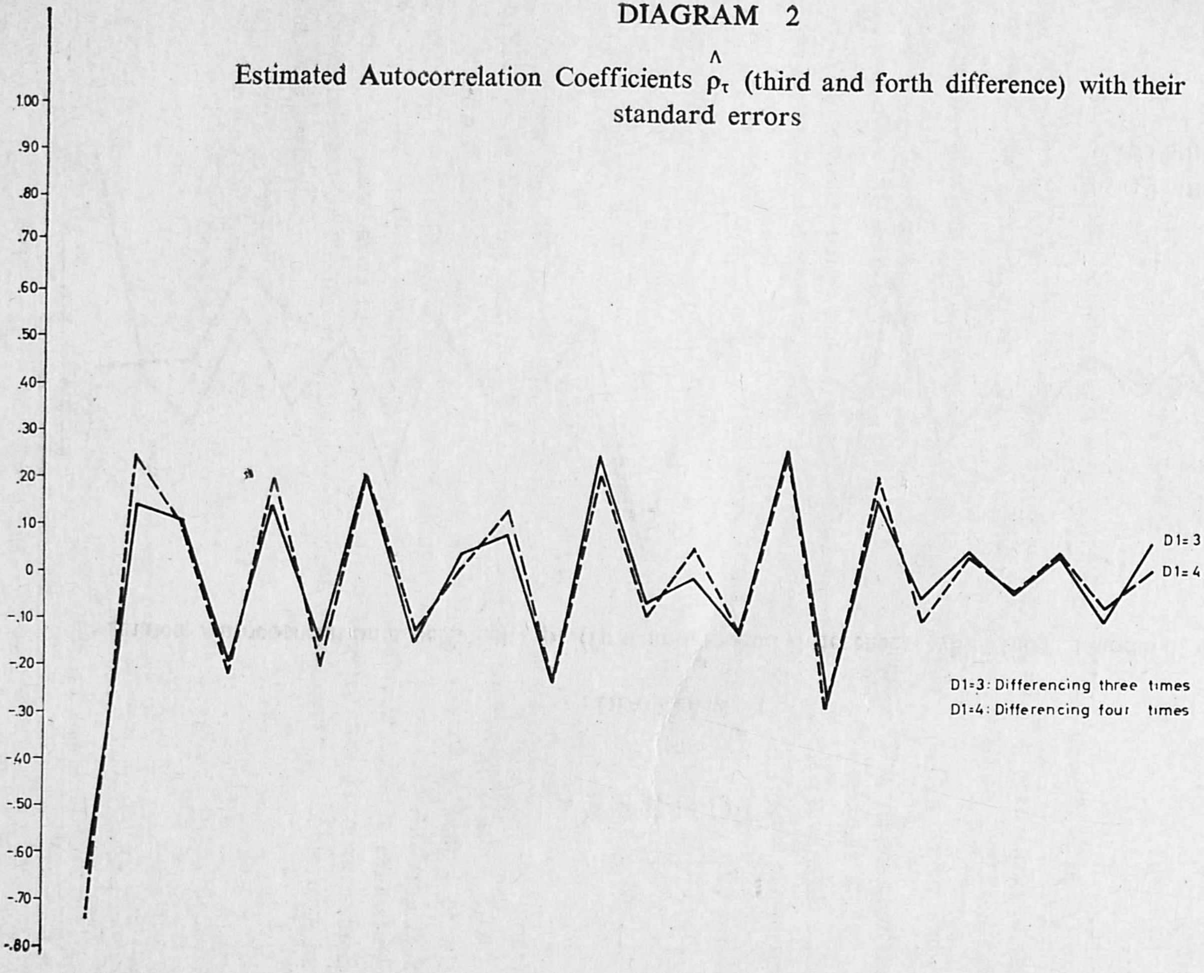
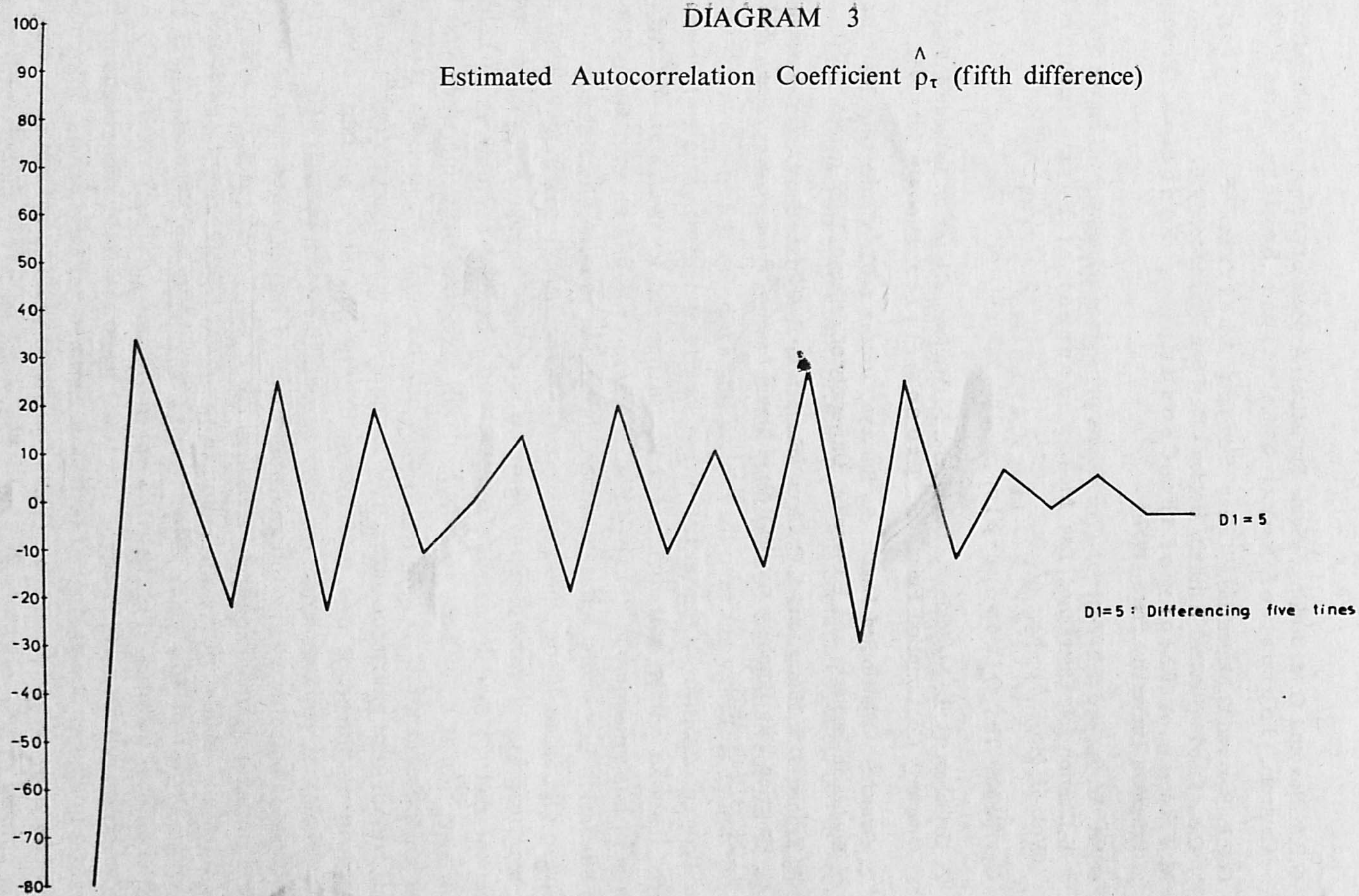


DIAGRAM 3

Estimated Autocorrelation Coefficient $\hat{\rho}_\tau$ (fifth difference)



REFERENSES

1. G.E.P. Box and G.M. Jenkins, «Some Statistical Aspects of Adaptive Optimization and Control», *Journal of Royal Statistical Society*, Series B, Vol. 24.

2. G.E.P. Box and G.M. Jenkins, *Time Series Analysis: Forecasting and Control* (Revised Edition), Holden - Day, San Francisco, 1976.

3. M. Friedman, *A Theory of the Consumption Function*, Princeton: Princeton University Press, 1957.

4. Fuller, W. A., ard J.E. Martin, «The Effects of Autocorrelated Errors on the Statistical Estimation of Distributed Lag Models», *Journal of Farm Economics*, Vol. 63, No 1, 71 - 82.

5. Gamaletsos Th., *Οἰκονομετρία*, pp. 239 - 247.

6. Z. Griliches, G. L. Maddala, R. Lucas, and N. Wallace, «Notes on Estimated Aggregate Quarterly Consumption Function», *Econometrica*, Vol. 30, July 1962.

7. Z. Griliches, «Distributed Lags : A. Survey», Report 6602, Center for Mathematical Studies in Business and Economics, University of Chicago, 1966, mimeo.

8. A. Zellner, D.S. Huang, and L.C. Chau, «Further Analysis of the Short - rum Consumption Function with Emphasis on the Role of Liquid Assets», *Econometrica*, Vol. 33, No 3, July 1965.