THE FITTING OF TIME-SERIES MODELS

by

James Durbin
University of North Carolina
and
University of London

This research was supported by the Office of Naval Research under Contract No. Nonr-855(09) for research in probability and statistics at Chapel Hill. Reproduction in whole or in part for any purpose of the United States Government is permitted.

Institute of Statistics
Mimeograph Series No. 244
December, 1959
THE FITTING OF TIME-SERIES MODELS 1,2

by

J. Durbin

University of North Carolina
and London School of Economics

1. Introduction

The purpose of this paper is to review methods of efficient estimation of the parameters in some of the models commonly employed in time-series analysis. The models we shall consider are the following:

The autoregressive model

(1) $u_t + a_1 u_{t-1} + \ldots + a_k u_{t-k} = \epsilon_t \quad (t = 1, \ldots, n)$;

Regression on fixed $X$'s and lagged $y$'s

(2) $y_t + a_1 y_{t-1} + \ldots + a_p y_{t-p} = \beta_1 x_{1t} + \ldots + \beta_q x_{qt} + \epsilon_t \quad (t = 1, \ldots, n)$;

Regression on fixed $X$'s with autoregressive disturbances

(3) $y_t = \beta_1 x_{1t} + \ldots + \beta_q x_{qt} + u_t$ \hspace{1cm} \text{where} \hspace{1cm} $u_t + a_1 u_{t-1} + \ldots + a_p u_{t-p} = \epsilon_t \quad (t = 1, \ldots, n)$;

The moving-average model

(4) $u_t = \epsilon_t + \beta_1 \epsilon_{t-1} + \ldots + \beta_h \epsilon_{t-h} \quad (t = 1, \ldots, n)$;

---

1 This research was supported by the Office of Naval Research under Contract No. Nonr-855(09) for research in probability and statistics at Chapel Hill. Reproduction in whole or in part for any purpose of the United States Government is permitted.

2. The autoregressive model

There is an underlying unity in the methods of estimation to be discussed in this paper arising from the fact that they all depend fundamentally on the method of least squares. For the autoregressive model

\[ u_t + a_1 u_{t-1} + \cdots + a_k u_{t-k} = \varepsilon_t \]
the minimisation of $\frac{1}{n} \sum_{t=1}^{n} (u_t + a_1 u_{t-1} + \cdots + a_k u_{t-k})^2$ leads directly
to estimates $a_1, \ldots, a_k$ which are efficient and easy to compute.

The sampling properties of $a_1, \ldots, a_k$ were first investigated by
Mann and Wald [8] who showed that they are the same asymptotically
as those of least-squares estimates of regression coefficients in
multivariate normal systems. (Actually Mann and Wald's model con-
tained a constant $a_0$ and therefore differs slightly from (1) but this
does not substantially affect the conclusions).

It is sometimes preferable to work with the asymptotically equi-
valent values $a_1', \ldots, a_k'$ obtained from the equations

$$
\begin{align*}
    r_1 + a_1' + r_1 a_2' + \cdots + r_{k-1} a_k' &= 0 \\
    r_2 + r_1 a_1' + a_2' + \cdots + r_{k-2} a_k' &= 0 \\
    \vdots \\
    r_k + r_{k-1} a_1' + \cdots + a_k' &= 0
\end{align*}
$$

(6)

where $r_i$ is the $i^{th}$ sample serial correlation coefficient. If
desired the $r_i$'s can be replaced by estimates of the serial covariances.

For all except small values of $k$ the equations (6) are easier to solve
than the least-squares equations owing to their possession of a more
symmetric structure. It is found that a pivotal reduction of (6)
reduces to the recurrence relations

$$
a_{ss} = \frac{r_s + a_{s-1} r_s + a_{s-2} + \cdots + a_{s-1,s-1}}{1 + a_{s-1} r_{s-1} + \cdots + a_{s-1,s-1} r_{s-1}}
$$

(7) 

$(s = 1, \ldots, k)$
(8) \[ a_{sr} = a_{s-l,r} + a_{ss}a_{s-l,s-r} \quad (r = 1, \ldots, s-1), \]

using \( a_{11} = -1 \) as the starting value. The quantities \( a_{s1}, \ldots, a_{ss} \) are the coefficients of the best-fitting autoregressive model of order \( s \), while \(-a_{22}, \ldots, -a_{kk}\) are estimates of the partial correlation coefficients between observations \( 2, \ldots, k \) time periods apart with intermediate observations held fixed. Apart from yielding this information of incidental interest, the use of (7) and (8) is decidedly more expeditious than a direct solution of (6). The final coefficients \( a_{k1}, \ldots, a_{kk} \) are identically equal to the values \( a_{1}', \ldots, a_{k}' \), obtained from (6).

3. Regression on fixed x's and lagged y's

The extension of the autoregressive model to include fixed x's appears to have been first considered about 1945 by Cowles Commission writers in connection with the study of simultaneous regression systems (see Koopmans et al. [7]). In this paper we are concerned only with the single-equation model

\[ y_t = \alpha_1 y_{t-1} + \cdots + \alpha_p y_{t-p} = \beta_1 x_{1t} + \cdots + \beta_q x_{qt} + \epsilon_t \quad (t = 1, \ldots, n) \]

where \( \{x_{1t}\}, \ldots, \{x_{qt}\} \) are sequences of constants. Putting \( y_{t-1} = x_{q+i,t} \) and \( \alpha_i = -\beta_{q+i} \) (\( i = 1, \ldots, p \)) the model can be written in the form

\[ y_t = \beta_1 x_{1t} + \cdots + \beta_q x_{q+p} x_{q+p,t} + \epsilon_t \quad (t = 1, \ldots, n), \]

or in an obvious matrix notation,

\[ y = X\beta + \epsilon, \]

where \( X \) is a \( nx(p+q) \) matrix.
The least-squares estimators of $\beta_1, \ldots, \beta_{q+p}$ are the elements of the vector $b = (X'X)^{-1}X'y$. If $X$ were a matrix of constants, least-squares theory would tell us that the vector of discrepancies $b - \beta$ has vector mean zero and variance matrix $\sigma^2(X'X)^{-1}$. For the present model these results do not hold since some of the elements of $X$ are random variables. However we can obtain analogous results by introducing the matrix $t = \left[ E(X'X) \right]^{-1} X'X$, where $E(X'X)$ denotes the matrix whose elements are the expected values of the corresponding elements of $X'X$. The matrix $t$ will usually converge stochastically to the unit matrix as $n \to \infty$.

It is shown in [41] that $t(b - \beta)$ has vector mean zero and variance matrix $\sigma^2 E(X'X)$. It is also shown that when the $\varepsilon$'s are normal this variance matrix is minimal in a certain sense. Letting

$$s^2 = \frac{1}{n-\varepsilon} \sum_{t=1}^{n} \left( y_t - b_1 x_{1t} \cdots \cdots b_{q+p} t x_{q+p,t} \right)^2$$

it is shown further that under certain assumptions $E(s^2) = \sigma^2 + O(\frac{1}{n})$ and that $b_1, \ldots, b_k$ are asymptotically multinormal. The implication of these results is that least-squares theory applies asymptotically to the model (2).

4. Regression on fixed $x$'s with autoregressive errors

Of greater relevance in many investigations is the model

$$y_t = \beta_1 x_{1t} \cdots + \beta_q x_{qt} + u_t \quad (t = 1, \ldots, n),$$

where the $u_t$'s are autocorrelated. It is well known that a simple least-squares analysis of data from such a model can be seriously misleading.
owing to the inefficiency of the least-squares estimators of the
\( \beta \)'s and to the biasedness of their estimates of variance (see for
example the discussion by Cochrane and Orcutt [3], Watson [10]
and Anderson [1]). It is true that for certain special cases, in­
cluding regressions on polynomial trends and seasonal constants, the
least-squares coefficients have been shown to be asymptotically ef­
ficient (Grenander and Rosenblatt [6] and Anderson and Anderson [2]).
Nevertheless the least-squares estimator of variance remains biased
and the use of analysis-of-variance methods for testing hypotheses
and setting confidence limits can be expected to give incorrect re­s­
ults.

It is often reasonable to assume that the \( u_t \)'s have the auto­
regressive structure

\[
(11) \quad u_t + a_1 u_{t-1} + \cdots + a_p u_{t-p} = \varepsilon_t .
\]

(10) may then be transformed to the form

\[
(12) \quad y_t + a_1 y_{t-1} + \cdots + a_p y_{t-p} = \beta_1 x_{1t} + a_1 \beta_1 x_{1t-1} + \cdots + a_p \beta_1 x_{1t-p} + \varepsilon_t .
\]

This now has the same structure as (2) except that relations exist
between the coefficients. Conceptually the simplest approach to the
estimation problem would be to minimize \( \sum_{t=1}^{n} \varepsilon_t^2 \), when this is expressed
in terms of the \( y \)'s and \( x \)'s, with respect to the \( a \)'s and \( \beta \)'s. How­
ever, since some of the coefficients in (12) are quadratic in the
unknowns this procedure results in non-linear estimating equations
which are usually unmanageable for practical use.
Before going on to discuss general methods we draw attention to an important special case in which simple methods based on least-squares theory do give satisfactory results. This arises when each \(x_{i,t-j}\) (\(i=1, \ldots, q; j=1, \ldots, p\)) can be expressed as a linear function of \(x_{1t}, \ldots, x_{qt}\). Examples are polynomial trends, seasonal constants and periodic regressions. (12) then reduces to the form (2) with functionally independent coefficients which can be legitimately estimated by least squares. We illustrate the procedure by considering the case of regression on a pure periodic function with first-order autoregressive disturbances, i.e.

\[y_t = \beta_1 \cos \lambda t + \beta_2 \sin \lambda t + u_t,\]

where \(u_t + au_{t-1} = \varepsilon_t\) and where \(\lambda\) is known.

We have

\[y_t + ay_{t-1} = \beta_1 \cos \lambda t + a\beta_1 \cos(\lambda(t-1)) + \beta_2 \sin \lambda t + a\beta_2 \sin(\lambda(t-1)) + \varepsilon_t\]

\[= \gamma_1 \cos \lambda t + \gamma_2 \sin \lambda t + \varepsilon_t,\]

where \(\gamma_1 = \beta_1 + a\beta_1 \cos \lambda - a\beta_2 \sin \lambda,\)

\(\gamma_2 = \beta_2 + a\beta_1 \sin \lambda + a\beta_2 \cos \lambda,\)

Efficient estimates \(a, c_1\) and \(c_2\) are then obtained by minimising

\[\sum_{t=1}^{n} (y_t + ay_{t-1} - \gamma_1 \cos \lambda t + \gamma_2 \sin \lambda t)^2,\]

whence estimates \(b_1, b_2\) of \(\beta_1, \beta_2\) result from the equations

\[(1 + a \cos \lambda) b_1 - a \sin \lambda b_2 = c_1\]

\[a \sin \lambda b_1 + (1 + a \cos \lambda) b_2 = c_2.\]
\( \sigma^2 \) is estimated by \( s^2 = \frac{1}{n-3} \sum_{t=1}^{n} (y_t + ay_{t-1} - c_1 \cos \lambda t - c_2 \sin \lambda t)^2 \).

For simplicity of exposition our discussion of the general problem will be based mainly on the two-coefficient model

\[(13) \quad y_t = \beta x_t + u_t, \quad \text{where} \]

\[(14) \quad u_t + \alpha u_{t-1} = \varepsilon_t, \quad (t=1, \ldots, n). \]

It was pointed out by Cochrane and Orcutt \([3]\) that if \( \alpha \) were known we could employ an autoregressive transformation \( y_t' = y_t + \alpha y_{t-1} \)

\[x_t' = x_t + \alpha x_{t-1}\]

to put the model in the form

\[y_t' = \beta x'_t + \varepsilon_t,\]

to which least squares can be applied quite validly. For the case of unknown \( \alpha \) they suggested that one should insert in the autoregressive transformation either a value of \( \alpha \) guessed on a priori grounds or a value estimated from the residuals of a fitted least-squares regression, further iterations being carried out if desired. The first of these suggestions is computationally attractive, though inefficient, while the second, though efficient, is computationally burdensome.

An approach will now be outlined which leads to estimates which are efficient and which are not too onerous to compute. From \((13)\) and \((14)\) we have

\[(15) \quad y_t + ay_{t-1} = \beta x_t + \gamma x_{t-1} + \varepsilon_t, \]

where \( \gamma = c\beta \). If we were to ignore the restriction \( \gamma = c\beta \) and regard \( \gamma \) as a free parameter, \((15)\) would have the form \((2)\) so that the least-squares estimators \( a, b, c \) obtained by minimising \( \sum_{t=1}^{n} (y_t + ay_{t-1} - bx_t - cx_{t-1})^2 \)
would be efficient estimators of $\alpha$, $\beta$, $\gamma$. To obtain efficient estimators of $\alpha$ and $\beta$ we need only therefore consider the joint distribution of $a$, $b$, $c$.

Now $y_t + ay_{t-1} - bx_t - ox_{t-1} = y_t + au_{t-1} - bx_t - (c - \alpha \beta) x_{t-1}$
since $y_{t-1} = \beta x_{t-1} + u_{t-1}$. Consequently $a$, $b$ and $c - \alpha \beta$ are the least-squares coefficients of regression of $y_t$ on $-u_{t-1}$, $x_t$ and $x_{t-1}$.

The corresponding true coefficients are $\alpha$, $\beta$ and zero in virtue of the relation

$$ (16) \quad y_t + au_{t-1} = \beta x_t + \epsilon_t. $$

By a slight extension of the results of the previous section we know that least-squares regression theory applies asymptotically to (16). Consequently the quantities $a - \alpha$, $b - \beta$ and $c - \alpha \beta$ are asymptotically normally distributed with zero means and variance matrix $\sigma^2 A^{-1}$, where $A$ is the expected value of the matrix

$$
\begin{bmatrix}
\Sigma u_{t-1}^2 & \Sigma u_{t-1}x_t & \Sigma u_{t-1}x_{t-1} \\
\Sigma u_{t-1}x_t & \Sigma x_t^2 & \Sigma x_t x_{t-1} \\
\Sigma u_{t-1}x_{t-1} & \Sigma x_t x_{t-1} & \Sigma x_{t-1}^2
\end{bmatrix}.
$$

Since $E(\Sigma u_{t-1}x_t) = E(\Sigma u_{t-1}x_{t-1}) = 0$ it follows that the asymptotic distribution of $a$, $b$ and $c - \alpha \beta$ has a density which is the limit as $n \to \infty$ of the expression
On maximising the exponent of (17) with respect to $\alpha$ and $\beta$ we find that their efficient estimates are

\[ \hat{\alpha} = a \]
\[ \hat{\beta} = \frac{\sum (x_t + a x_{t-1}) (y_t + a y_{t-1})}{\sum (x_t + a x_{t-1})^2} \]

It is remarkable that $\hat{\beta}$ is precisely the same estimator as is obtained by using $a$ as an estimator of $\alpha$ in an autoregressive transformation. The same procedure was arrived at earlier by the author [4] using a rather different approach.

Ignoring the difference between $\frac{1}{n} \sum x_t^2$ and $\frac{1}{n} \sum x_{t-1}^2$, as is legitimate to the order of accuracy considered here, we find that (18) reduces to

\[ \hat{\beta} = \frac{(1 + ar)b + (a + r)c}{1 + 2ar + a^2} \]

where $r = \frac{\sum x_t x_{t-1}}{\sum x_t^2}$. Note that $a$ and $\hat{\beta}$ are asymptotically independently distributed.

The treatment of the general model (3) follows along similar lines. We shall confine ourselves here to the presentation of a brief summary of the computing routine, referring the reader to [4] for further theoretical discussion. For simplicity of exposition let us suppose that the variables $x_{1t}$, $x_{1,t-1}$, ..., $x_{qt}$, ..., $x_{q,t-p}$ are linearly
independent. (The outstanding case to the contrary is the common
one in which the model contains a constant term, i.e. \( x_{it} \) equals
unity for some \( i \) and all \( t \); this, however, is easily dealt with by
working throughout with deviations from sample means as in ordinary
regression analysis. Modifications for other cases are easily worked
out ad hoc.

Suppose that the normal equations for the least-squares fitting
of the regression of \( y_t \) on \( x_{it}, x_{it-1}, \ldots, x_{it-p}, \ldots, x_{qt}, \ldots \),
\( x_{q,t-p}, \ldots, y_{t-1}, \ldots, y_{t-p} \), the variables being taken in this order,
are denoted by

\[
A_1 b = c_1,
\]

where \( A_1 \) is the \((p + q + pq) \times (p + q + pq)\) matrix of sums of squares
and products of the variables \( x_{it}, \ldots, x_{qt}, y_{t-1}, \ldots, y_{t-p} \),
where \( c_1 \) is the vector of sums of products of these variables with
\( y_t \), and where \( b \) is the vector of regression coefficients. If the
equations are solved by a method such as the abbreviated Doolittle
method note that it is only necessary to carry the back solution
far enough to give the coefficients \( a_1, \ldots, a_p \) of \( -y_{t-1}, \ldots, -y_{t-p} \).

Form a new matrix \( A_2 \) whose first row is obtained by multiplying
the first \( p+1 \) rows of \( A_1 \) by \( 1, a_1, \ldots, a_p \) respectively and adding,
whose second row is obtained by taking the second group of \( p+1 \) rows
of \( A_1 \), multiplying by \( 1, a_1, \ldots, a_p \) respectively and adding. Con­
tinue in this way until \( A_2 \) has \( q \) rows. Repeat the process on \( c_1 
\)
to give a new vector \( c_2 \) containing \( q \) elements. Repeat the process on the
columns of $A_2$ to give a new matrix $A_3$ with $q$ rows and columns. Let $c_3$ be the vector obtained by subtracting from $c_2$ the sum of $a_1$ times the last column of $A_2$, $a_2$ times the second last column, $\ldots$, $a_p$ times the $p^{th}$ column of $A_2$ counting backwards from the last column. Then the solution $\hat{\beta}$ of the equation

$$A_3 \hat{\beta} = c_3$$

is the vector of efficient estimators of $\beta_1, \dotsc, \beta_q$. Its estimated variance matrix is $s^2 A_3^{-1}$, where

$$s^2 = \frac{1}{n-p-q} \sum_{t=1}^{n} \left( y_t + a_1 y_{t-1} + \cdots + a_p y_{t-p} \right)^2 \hat{\beta}^T c_3.$$ 

5. **The moving-average model**

The special problems of fitting moving-average models can be appreciated from a consideration of the first-order model

$$u_t = \varepsilon_t + \beta \varepsilon_{t-1} \quad (t = 1, \ldots, n).$$

A simple estimator of $\beta$ can be obtained by equating the theoretical value of the first serial correlation, namely $\beta/(1+\beta^2)$, to the sample value $r_1$. However, the estimator was shown to be inefficient by Whittle [11,12] who proposed the use of an approximate maximum-likelihood estimator equivalent to the solution to the equation

$$\frac{\beta}{\gamma^2} \left[ \frac{1}{1-\beta^2} (1 - 2\beta r_1 + 2\beta^2 r_2 - \cdots) \right] = 0$$

where $r_i$ is the $i^{th}$ sample serial correlation.

Although efficient, this estimator is difficult to calculate and the method does not easily extend to higher-order models. The
author [5] has therefore suggested a different method in which a
\( k^{th} \)-order autoregressive model is first fitted to the data, \( k \)
being taken to be large. It is shown in [5] that the fitted coeffi­
cients \( a_1, \ldots, a_k \) have an asymptotic distribution with the approximate
density

\[
\text{constant} \times (1-\beta^2)^{-1/2} \exp \left[ -\frac{n}{2} \sum_{i=0}^{k-1} (a_{i+1} + \beta a_i)^2 + \beta^2 a_k^2 \right].
\]

Neglecting the factor \( (1-\beta^2)^{-1/2} \) since this is of small order in
\( n \) compared with the remainder and maximising the exponent with respect
to \( \beta \) we obtain as our estimator of \( \beta \),

\[
b = - \frac{\sum_{i=0}^{k-1} a_i a_{i+1}}{\sum_{i=0}^{k-2} a_i^2} \quad (a_0 = 0),
\]

the efficiency of which can be made as close to unity as desired by
taking \( k \) sufficiently large.

For the general model

\[
u_t = \varepsilon_t + \beta_1 \varepsilon_{t-1} + \ldots + \beta_h \varepsilon_{t-h} \quad (t = 1, \ldots, n)
\]

the same approach yields estimators \( b_1, \ldots, b_h \) which are obtained
as the solution to the linear equations

\[
\sum_{i=1}^{h} \sum_{j=1}^{r-j} A_{i+j} b_j = -A_r \quad (r = 1, \ldots, h),
\]

where \( A_r = \sum_{i=0}^{k-r} a_i a_{i+r} \), the \( a_i \)'s being as before.

6. The autoregressive model with moving-average errors.
This model, which has the generating equation

\[(24) \quad u_t + \gamma_1 u_{t-1} + \cdots + \gamma_p u_{t-p} = \xi_t + \delta_q \xi_{t-q} \quad (t=1, \ldots, n),\]

has greater theoretical importance than the attention paid to it in time-series literature would appear to indicate. Firstly, it is the general model of which the autoregressive and moving-average models are special cases. Secondly, when \(q=p-1\) it is the only one of the three models whose structure is invariant under changes in the time-period between successive observations, a fact pointed out by Quenouille [9]. Thirdly, equi-spaced observations from a continuous stochastic process generated by a linear stochastic differential equation, or having a rational spectral density, conform to a discrete model (24) with \(q = p-1\). Consequently a solution to the problem of efficient fitting of (24) also gives as a by-product the solution to the problems of fitting stochastic differential equation models and of estimating rational spectral densities from discrete data. Yet in spite of the theoretical importance of the model only Quenouille [9] appears to have considered the fitting problem; however Quenouille did not attempt to discuss efficient methods of estimation.

Two methods of fitting will now be described. The first is non-iterative but is not fully efficient. The second is an iterative method in which the autoregressive and moving-average parameters are estimated alternately. It is hoped to investigate the performance
of both methods by means of sampling experiments and to publish
the results later.

Let us begin by considering the first-order model

\[ u_t + \gamma u_{t-1} = \xi_t + \delta \varepsilon_{t-1} \quad (t=1, \ldots, n). \]

Let \( a_1, \ldots, a_k \) denote the coefficients of a fitted autoregressive
model of large order \( k \) and let \( \varepsilon_t \) denote the residual \( u_t + a_1 u_{t-1} \)
\[ + \cdots + a_k u_{t-k} \]. In the first method of estimation we replace \( \xi_{t-1} \)
in \( (25) \) by \( \varepsilon_{t-1} \) and estimate \( \gamma \) and \( \delta \) by the values of \( c \) and of \( d \)
which minimise \( \sum (u_t + c u_{t-1} - d \varepsilon_{t-1})^2 \). This leads to the equations

\[ c \sum u_{t-1}^2 - d \sum u_{t-1} \varepsilon_{t-1} = -\sum u_t u_{t-1} \]
\[ c \sum u_{t-1} \varepsilon_{t-1} - d \sum \varepsilon_{t-1}^2 = -\sum u_t \varepsilon_{t-1}, \]

the solution of which is asymptotically equivalent to the expressions

\[ c = - \frac{a_1 r_2 + a_2 r_3 + \cdots + a_k r_{k+1}}{a_1 r_1 + a_2 r_2 + \cdots + a_k r_k}, \]
\[ d = c + \frac{r_1 + a_1 r_2 + \cdots + a_k r_{k+1}}{1 + a_1 r_1 + \cdots + a_k r_k}. \]

The method is readily extended to cover higher-order systems and can
be used to give starting values for the second method, which we now
describe.

First let us consider the estimation of \( \delta \) for a given value of \( \gamma \).
Suppose that the true values of the first \( k \) autoregressive coefficients
are \(a_1, \ldots, a_k\), the fitted values being denoted by \(a_1, \ldots, a_k\) as before. Using Mann and Wald's results \([8]\) we know that for large \(k\) the asymptotic distribution of \(a_1, \ldots, a_k\) is normal with density

\[
(28) \quad \text{constant } x \exp \left[ -\frac{n}{2\sigma^2} \sum_{i,j=1}^{k} (a_i - \alpha_i) (a_j - \alpha_j) E(u_{t-i} u_{t-j}) \right].
\]

The quadratic form in the exponent can be represented operationally as

\[
(29) \quad Q = \mathcal{E} \left( \sum_{i=1}^{k} (a_i - \alpha_i) u_{t-i} \right)^2
\]

where \(\mathcal{E}\) denotes the operation of taking an expectation over variation of the \(u_i\)'s, the \(a_i\)'s being regarded as fixed constants.

Suppose that the true value of \(\gamma\) were known and the following transformation from \(a_1, \ldots, a_k\) and \(\alpha_1, \ldots, \alpha_k\) to \(\ell_1, \ldots, \ell_k\) and \(\lambda_1, \ldots, \lambda_k\) were made,

\[
\begin{align*}
\ell_1 &= a_1 + \gamma \\
\ell_2 &= a_2 + \gamma \ell_1 \\
& \quad \vdots \\
\ell_k &= a_k + \gamma \ell_{k-1}
\end{align*}
\]

Substituting in (29) we have

\[
(30) \quad Q = \mathcal{E} \left( (\ell_1 - \lambda_1) (u_{t-1} + \gamma u_{t-2}) + \cdots + (\ell_{k-1} - \lambda_{k-1}) (u_{t-k+1} + \gamma u_{t-k}) + (\ell_k - \lambda_k) u_{t-k} \right)^2.
\]
Now $u_t + \lambda u_{t-1} = \epsilon_t + \delta \epsilon_{t-1}$. Consequently, on putting $z_t = u_t + \gamma u_{t-1}$ we see that $z_t$ satisfies the moving-average process $z_t = \epsilon_t + \delta \epsilon_{t-1}$.

Moreover $u_{t-k} = z_{t-k} - \gamma z_{t-k-1} + \gamma^2 z_{t-k-2} + \cdots$. Consequently on putting $q_{k+r} = (-\gamma)^r q_k$ and $\lambda_{k+r} = (-\gamma)^r \lambda_k (r = 1, 2, \ldots)$, (30) gives

$$Q = \sum_{i=1}^{\infty} (q_i - \lambda_i) z_{t-i}^2$$

From the fact that the true autoregressive coefficients are generated by the relation $1 + \alpha_1 z + \alpha_2 z^2 + \cdots = (1 + \gamma z)(1 + \delta z)^{-1}$, it follows that for large $k$, $\lambda_i = (\alpha_i)^i (i = 1, \ldots, k)$ as accurately as desired. Since $\lambda_k$ can be made arbitrarily small the error committed by taking $\lambda_{k+r} = (-\delta)^r \lambda_k$ in (30) in place of $\lambda_{k+r} = (-\gamma)^r \lambda_k (r = 1, 2, \ldots)$ can be made arbitrarily small. Thus to a high degree of accuracy (31) holds with $z_{t-1}, z_{t-2}, \ldots$ corresponding to a moving-average model for which $\lambda_1, \lambda_2, \ldots$ are the true autoregressive coefficients.

Comparing (31) with (29) we see that $\ell_1, \ell_2, \ldots$ behave like autoregressive coefficients fitted to data corresponding to the model

$$z_t = \epsilon_t + \delta \epsilon_{t-1} \quad (t=1, \ldots, n)$$

Consequently it follows from (22) that the efficient estimator of
In practice it will probably suffice to terminate the summations in this expression at about $i = k$. Note that we need not take explicit account of the fact that the "constant" in (28) depends on the covariance determinant of $a_1, \ldots, a_k$, which in turn depends on the unknown parameter $\delta$, since the determinant is of small order in $n$ compared with the exponent of (28).

Let us now consider the converse problem, i.e. the estimation of $\gamma$ given the true value of $\delta$. Define $w_1, \ldots, w_n$ by the relation

$$u_t = w_t + \delta w_{t-1} \quad (t = 1, \ldots, n)$$

where $w_0$ is either defined arbitrarily or taken equal to $u_0 - \delta u_{-1} + \delta^2 u_{-2} \ldots \ldots$

The $w$'s then satisfy the autoregressive model

$$w_t + \gamma w_{t-1} = \epsilon_t$$

Consequently $\gamma$ is efficiently estimated by the expression

$$c = - \frac{E' w_t w_{t+1}}{\sum w_t^2}$$

where $E'$ denotes summation over the range of possible values of $t$. 

\[ d = - \sum_{i=0}^{\infty} \frac{\ell_i \ell_{i+1}}{\sum_{i=0}^{\infty} \ell_i^2} \quad (\ell_0 = 1). \]
divided by the number of terms summed.

Let

\[ s_r = \sum u_t u_{t+r} \]
\[ p_r = \sum u_t w_{t+r} \]
\[ q_r = \sum w_t w_{t+r} \]

It is easy to verify that to a good approximation we have the relations

\[(34)\] \[ p_r = s_r - \delta p_{r-1} \quad (r = -k+1, -k+2, \ldots, k) \]

\[(35)\] \[ q_r = p_r - \delta q_{r+1} \quad (r = k-1, k-2, \ldots, 0) . \]

Applying (34) recursively taking \( p_{-k} = s_{-k} \), and then (35) recursively taking \( q_k = p_k \), we obtain \( q_1 \) and \( q_0 \) from which we obtain the estimate of \( \gamma \) as

\[(36)\] \[ c = \frac{q_1}{q_0} . \]

By applying (32) and (36) alternately we obtain an iterative method of estimating \( \gamma \) and \( \delta \). (26) or (27) can be used to provide a starting point.

The treatment of the general model

\[ u_t + \gamma_1 u_{t-1} + \cdots + \gamma_p u_{t-p} = \varepsilon_t + \delta_1 \varepsilon_{t-1} + \cdots + \delta_q \varepsilon_{t-q} \quad (t=1, \ldots, n) \]

follows similar lines. The fitted autoregressive constants \( a_1, \ldots, a_k \) are transformed to \( l_1, \ldots, l_k \) by the relations
\[ a_1 = \ell_1 + \gamma_1 \]
\[ a_2 = \ell_2 + \gamma_1 \ell_1 + \gamma_2 \]
\[ \vdots \]
\[ a_k = \ell_k + \gamma_1 \ell_{k-1} + \cdots + \gamma_p \ell_{k-p} \]

and further fits are obtainable from the expression
\[ \ell_r + \gamma_1 \ell_{r-1} + \cdots + \gamma_p \ell_{r-p} = 0 \quad (r = k+1, k+2, \ldots). \]

The fits behave approximately like autoregressive coefficients fitted to data generated by the moving-average model
\[ z_t = \epsilon_t + \delta_1 \epsilon_{t-1} + \cdots + \delta_q \epsilon_{t-q} \]

\( \delta_1, \ldots, \delta_q \) can therefore be estimated from equations (23) taking
\[ A_r = \sum_{i=0}^{\infty} \ell_i l_{i+r} \quad (\ell_0 = 1). \] In practice the summation can probably be truncated at about \( i = k \).

To estimate \( \gamma_1, \ldots, \gamma_p \) for given \( \delta_1, \ldots, \delta_q \) we use in place of (34) and (35) the expressions
\[ p_r + \delta_1 p_{r-1} + \cdots + \delta_q p_{r-q} = s_r \quad (r = -k+q, \ldots, k) \]
\[ q_r + \delta_1 q_{r-1} + \cdots + \delta_q q_{r-q} = r_r \quad (r = k-q, \ldots, 0). \]

Estimates \( c_1, \ldots, c_p \) are then obtained from the equations
\[ q_0 c_1 + q_1 c_2 + \cdots + q_{p-1} c_p = -q_1 \]
\[ (39) \quad q_1 c_1 + q_0 c_2 + \cdots + q_{p-2} c_p = -q_2 \]
\[ \vdots \]
\[ q_{p-1} c_1 + \cdots + q_0 c_p = -q_p \]
References


