The Annals of Statistics
1975, Vol. 3, No. 6, 1354–1360

NONINVERTIBLE TRANSFER FUNCTIONS AND THEIR FORECASTS

BY DAVID A. PIERCE

Federal Reserve Board

A transfer function relating a time series y_t to present and past values of a series x_t need not possess an inverse. When (x_t, y_t) is a covariance stationary process, it is shown that noninvertibility in this transfer function has the effect of reducing the error variance of the minimum mean-square-error predictor of y_t one or more steps ahead. In deriving these results a "dual" series to x_t is constructed, which has univariate stochastic structure identical to that of x_t itself, and an associated dual transfer function relating it to y_t which is invertible.

1. Introduction. The model of this paper is that of a bivariate linear, nonsingular, purely nondeterministic covariance-stationary time series $\{(x_t, y_t): t = \cdots, -1, 0, 1, \cdots\}$ in which the cross correlation between y_t and x_{t+k} is zero for positive k. Specifically, the following assumptions are made:

A1. The variables x and y possess a relationship of the form

$$y_t = \sum_{j=0}^{\infty} \tau_j x_{t-j} + e_t = \tau(B)x_t + e_t ; \tag{1.1}$$

the *transfer function* $\tau(B) = \sum_{j=0}^{\infty} \tau_j B^j$ is a polynomial in the backshift operator B defined by $B^j x_t = x_{t-j}$, satisfying $\sum |\tau_j| < \infty$.

A2. The two time series $\{x_t\}$ and $\{e_t\}$ are independent, each representable in the form

$$x_t = \sum \xi_j w'_{t-j} = \xi(B)w_t' \tag{1.2}$$

and

$$e_t = \sum \psi_j a_{t-j} = \psi(B)a_t , \tag{1.3}$$

where w_t' and a_t are serially and mutually independent white noise sequences; $\xi(B) = \sum_0^\infty \xi_j B^j$, $\psi(B) = \sum_0^\infty \psi_j B^j$; and $\xi_0 = \psi_0 = 1$.

A3. The operators $\xi(B)$ and $\psi(B)$ are *invertible*; i.e., x_t and e_t possess autoregressive representations [1]

$$\mu(B)x_t = w_t' , \qquad \pi(B)e_t = a_t ,$$

where $\mu(B)$, $\pi(B)$ are absolutely convergent and

$$\mu(B)\xi(B) = \pi(B)\psi(B) = 1 .$$

A4. The transfer function $\tau(B)$ is a rational function of B, i.e.

$$\tau(B) = \frac{\omega(B)}{\delta(B)} , \tag{1.4}$$

Received January 1974; revised April 1975.

AMS 1970 *subject classifications*. Primary 62M20; Secondary 62M10.

Key words and phrases. Transfer-function models, dynamic models, distributed lag models, forecasting, invertibility (of transfer function models), prediction.

where $\omega(B)$ and $\delta(B)$ are polynomials of finite orders k and k'. The roots of the auxiliary equation $\delta(z) = 0$ lie outside the unit circle.

A5. The transfer function $\tau(B)$ is *strictly noninvertible*, i.e. $\omega(z)$ in (1.4) has a root inside the unit circle.

It is the last of these, A5, that is the particular focus of this paper. While for stability the operators ξ and ψ are generally required to possess inverses, no such necessary restrictions exist on $\tau(B)$ for either the definition or the empirical modelling [2] of (1.1); and some of the consequences of noninvertibility of the transfer function are the focus of this study. Note that $\tau(B)$ is invertible if and only if $|\beta_j| > 1$, $j = 1, \cdots, k$ in the factorization of $\omega(z)$,

$$\omega(z) = \omega_k \prod_{j=1}^{k} (\beta_j + z) . \tag{1.5}$$

Assumption A5 is that for some j, $|\beta_j| < 1$. The term "noninvertibility" actually refers only to $|\beta_j| \leqq 1$ for some j, whence the use of the word "strict" in A5. In practice, as noted in [3, pages 43–44], due to rounding errors $|\beta_j| = 1$ will not generally occur, and the use of "noninvertibility" in the sequel will refer to strict noninvertibility. If in fact $\min_{1 \leqq j \leqq k} |\beta_j| = 1$ then the transfer function might be regarded as "near-invertible," and the present treatment can be extended to include this case under the "invertible" heading.

Suppose, for the remainder of this section only, that the transfer function $\tau(B)$ *is* invertible. Substituting (1.2) and (1.3) into (1.1),

$$\begin{aligned} y_t &= \tau(B)\xi(B)w_t' + \psi(B)a_t \\ &= \chi'(B)w_t' + \psi(B)a_t \\ &= \chi(B)w_t + \psi(B)a_t \quad (\chi_0 = 1) \\ &= h_t + e_t , \end{aligned} \tag{1.6}$$

where the normalization $\chi_0 = 1$ is possible since $\chi'(B)$ and $\chi(B)$ are invertible whenever $\tau(B)$ is. Then, to forecast a future observation y_{n+p}, $p \geqq 1$, given $S_n = \{(x_t, y_t) : t \leqq n\}$, under the criterion of minimum mean square error the optimal predictor $\hat{y}_n(p)$ is the expectation of y_{n+p} conditional on S_n (see [3] for example). For the model (1.6) this predictor is simply

$$\hat{y}_n(p) = \chi(B)w_{n+p}^* + \psi(B)a_{n+p}^* = \hat{h}_n(p) + \hat{e}_n(p) , \tag{1.7}$$

where

$$\begin{aligned} w_t^* &= w_t , \quad t \leqq n \\ &= 0 , \quad t > n \end{aligned}$$

and similarly for a_t^*. Moreover, the MSE of this forecast, also referred to as the p-step prediction variance, is

$$V(p) = \sigma_w^2 \sum_{j=0}^{p-1} \chi_j^2 + \sigma_a^2 \sum_{j=0}^{p-1} \psi_j^2 . \tag{1.8}$$

The assumption (for this paragraph only) that $\tau(B)$ is invertible implies that the series $\{w_t\}$ constitutes the observable past-history innovations of the linear process

$$h_t = \chi(B)w_t , \tag{1.9}$$

since $\chi(B)$ is the product of two invertible operators. Consequently, the single-period prediction MSE,

$$V(1) = \sigma_w^2 + \sigma_a^2 , \tag{1.10}$$

is the sum of the two innovation variances, a result which will be seen not to hold when A5 is true.

In the next section a framework is developed for analyzing the present case where the transfer function $\tau(B)$ is noninvertible. The quantitative effect of noninvertibility on the prediction variance (1.8) and (1.10) is the subject of Section 3.

2. Explicit and implicit representations. Suppose, in accordance with A5, that $m \geqq 1$ zeroes of the numerator $\omega(B)$ of the transfer function (1.4) lie within the unit circle. Denoting these by $-\alpha_1, \cdots, -\alpha_m$, it follows that the quantity

$$\nu(B) = \nu_0 + \nu_1 B + \cdots + \nu_{m-1}B^{m-1} + B^m = \prod_{i=1}^m (\alpha_i + B) \tag{2.1}$$

is a factor of $\omega(B)$ (if any of the α_j are zero then the leading coefficients of $\nu(B)$ vanish). Let $\omega^*(B)$ and $Q(B)$ be defined by

$$\omega(B) = \omega^*(B)\nu(B)$$
$$Q(B) = [\omega^*(B)/\delta(B)]\xi(B)/\xi_0\omega_0^* .$$

Thus $Q(B)$ is the normalized ($Q_0 = 1$) product of the operator $\xi(B)$ characterizing the stochastic process $\{x_t\}$ and the remainder of the transfer function. Defining

$$\chi^{(E)}(B) = Q(B)\nu(B) , \tag{2.2}$$

it follows that the component $[\tau(B)x_t]$ of y_t is [compare (1.6)]

$$h_t = \nu(B)Q(B)w_t = \chi^{(E)}(B)w_t . \tag{2.3}$$

It is important to note that (2.3) does not define a linear process the way that (1.9) does; the existence of interior roots implies that $\{w_t\}$ in (2.3), while observable given $\{x_t\}$, are unobservable given only $\{h_t\}$. Instead, the innovations of the series $\{h_t\}$, say $\{v_t\}$, satisfy

$$h_t = \chi^{(I)}(B)v_t \tag{2.4}$$

where (a) $\chi^{(I)}(B)$ is invertible and (b) the autocovariance function or spectrum of (2.4) is identical to that of (2.3). The structure of h_t as a linear process is given by

LEMMA 1. *The operator $\chi^{(I)}(B)$ in (2.4) is*

$$\chi^{(I)}(B) = \eta(B)Q(B) , \tag{2.5}$$

where

$$\eta(B) = 1 + \eta_1 B + \cdots + \eta_m B^m = \prod_{i=1}^m (1 + \alpha_i B) , \tag{2.6}$$

except that if b of the α's in (2.1) are zero, then $\eta(B)$ is of degree $m - b$.

PROOF. The covariance generating function of h_t is given by

$$\sigma_w^2 Q(B)Q(F)\nu(B)\nu(F) = \sigma_v^2 Q(B)Q(F)\eta(B)\eta(F),$$

where $F = B^{-1}$ is the "forward shift" operator. It is easily seen that, since $|\alpha_i| < 1$,

$$(\alpha_i + B)(\alpha_i + F) = (1 + \alpha_i B)(1 + \alpha_i F),$$

so that

$$\nu(B)\nu(F) = \eta(B)\eta(F), \tag{2.7}$$

and hence

$$\sigma_w^2 = \sigma_v^2. \tag{2.8}$$

Since $|\alpha_i| < 1$, $\chi^{(I)}(B)$ in (2.6) is invertible; thus (2.4) defines a stationary linear process with observable past-history innovations v_t. □

We shall refer to (2.3) and (2.4) as respectively the *explicit* and *implicit* representations of h_t. It can be shown that the coefficient of B^i in $\nu(B)$, $0 \leqq i \leqq m$, is the coefficient of B^{m-i} in $\eta(B)$. These results can be summarized as

THEOREM 1. *Corresponding to the stationary noninvertible representation*

$$h_t = Q(B)\nu(B)w_t = \chi^{(E)}(B)w_t$$

of the component $h_t = [\omega(B)/\delta(B)]x_t$ *of the transfer function-noise model* (1.1), *there exists a unique invertible representation*

$$h_t = Q(B)\eta(B)v_t = \chi^{(I)}(B)v_t.$$

The autocovariance structures implied by the dual linear operators $\chi^{(I)}(B)$ *and* $\chi^{(E)}(B)$, *or equivalently by* $\eta(B)$ *and* $\nu(B)$, *are identical, as are the variances of the innovations* w_t *and* v_t.

In addition to the dual representations of h_t *given by* (2.3) *and* (2.4), *there exist*

a) *the dual invertible transfer functions* $\tau(B)$ *and*

$$\frac{\eta(B)\omega^*(B)}{\delta(B)} = \tau^{(I)}(B), \tag{2.9}$$

and

b) *the dual independent variable series* x_t *and*

$$x_t^{(I)} = [\tau^{(I)}(B)]^{-1}h_t = \xi(B)v_t. \tag{2.10}$$

As univariate stochastic processes, $\{x_t^{(I)}\}$ *and* $\{x_t\}$ *are indistinguishable.*

3. Reduction in prediction variance. As seen above, noninvertibility in a transfer function implies that two sets of innovations, $\{w_t\}$ and $\{v_t\}$, are recoverable from the system whereas only one would be otherwise; in other words the series x_t contains additional information pertinent to y_t not imbedded within the dual series $x_t^{(I)}$. A measure of this additional information is the reduction in the forecast variance.

For the noninvertible transfer function model containing (2.3) as a component, it is easily seen that the forecast variance is given by (1.8) with $\chi_j^{(E)}$ replacing χ_j; for the dual implicit model (2.4), the MSE is also given by (1.8), noting that $\sigma_v^2 = \sigma_w^2$, by replacing χ_j with $\chi_j^{(I)}$. Consequently, the difference in prediction variance as a result of noninvertibility in the transfer function $\tau(B)$, is, letting $\sigma^2 = \sigma_v^2 = \sigma_w^2$

$$\Delta(p) = \sigma^2 \textstyle\sum_{j=0}^{p-1} \{(\chi_j^{(I)})^2 - (\chi_j^{(E)})^2\} . \tag{3.1}$$

In justifying the title of this section it is now shown that (3.1) is positive for $p = 1$ and nonnegative for general p.

THEOREM 2. *The single-step prediction variance reduction is*

$$\Delta(1) = \sigma^2\{1 - [\textstyle\prod_{i=1}^{m} \alpha_i]^2\} > 0 . \tag{3.2}$$

PROOF. Recalling that $Q_0 = 1$, (2.2) and (2.5) imply

$$\Delta(1) = \eta_0^2\sigma^2 - \nu_0^2\sigma^2 = \sigma^2(1 - \nu_0^2)$$

which, since $\nu_0 = \prod \alpha_i$, is (3.2). Since each $|\alpha_i|$ is less than unity this quantity is positive. □

The quantity $\Delta(1)$ is particularly amenable to interpretation since it depends only on σ^2 and α, whereas in general (3.1) depends on other aspects of the bivariate process (x_t, y_t) as embodied in $Q(B)$. However, (3.1) is never negative.

THEOREM 3. *For all positive integers* p,

$$\Delta(p) \geqq 0 . \tag{3.3}$$

PROOF. First consider the case where $\alpha_1, \cdots, \alpha_m$ are all real, for which the result will be seen to follow by application of

LEMMA 2. *If, for any convergent operator* $H(B) = \sum_{j=0}^{\infty} H_j B^j$,

$$[H(B)]_p = \textstyle\sum_{j=0}^{p} H_j^2 , \tag{3.4}$$

then for any $|\alpha| < 1$,

$$[(\alpha + B)H(B)]_p \leqq [(1 + \alpha B)H(B)]_p . \tag{3.5}$$

PROOF OF LEMMA. Since

$$(\alpha + B)H(B) = H_0\alpha + (H_0 + H_1\alpha)B + \cdots \tag{3.6}$$

and

$$(1 + \alpha B)H(B) = H_0 + (\alpha H_0 + H_1)B + \cdots , \tag{3.7}$$

it follows that the left hand side of (3.5) is

$$\alpha^2 \textstyle\sum_0^p H_j^2 + 2\alpha \sum_0^{p-1} H_j H_{j+1} + \sum_0^{p-1} H_j^2 \tag{3.8}$$

whilst the right hand side is

$$\textstyle\sum_0^p H_j^2 + 2\alpha \sum_0^{p-1} H_j H_{j+1} + \alpha^2 \sum_0^{p-1} H_j^2 \tag{3.9}$$

whence the right minus the left is

$$H_p^2(1 - \alpha^2) \geqq 0 . \tag{3.10}$$

□

To establish the theorem (for real roots), apply the lemma m times: first with $H(B) = Q(B) \prod_{i=2}^m (1 + \alpha_i B)$, $\alpha = \alpha_1$; then with $H(B) = Q(B)(\alpha_1 + B) \prod_{i=3}^m (1 + \alpha_i B)$, $\alpha = \alpha_2$; then with $H(B) = Q(B) \prod_{j=1}^2 (\alpha_j + B) \prod_{i=4}^m (1 + \alpha_i B)$, $\alpha = \alpha_3$; $\cdots$; and finally with $H(B) = Q(B) \prod_{j=1}^{m-1} (\alpha_j + B)$ and $\alpha = \alpha_m$. This gives m inequalities of the form (3.5), successive ones of which relate transitively; the left hand side of the last is thus not greater than the right hand side of the first, viz.

$$[\textstyle\prod_{i=1}^m (\alpha_i + B)Q(B)]_p \leqq [\prod_{i=1}^m (1 + \alpha_i B)Q(B)]_p ,$$

which replacing p by $p - 1$ is equivalent to $\Delta(p) \geqq 0$. □

Complex roots are handled in pairs according to

LEMMA 3. *If $\bar{\alpha}$ is the conjugate of α,*

$$[(\alpha + B)(\bar{\alpha} + B)H(B)]_p \leqq [(1 + \alpha B)(1 + \bar{\alpha} B)H(B)]_p . \tag{3.11}$$

PROOF OF LEMMA. If $(\alpha + B)(\bar{\alpha} + B) = \nu_0 + \nu_1 B + B^2$, then

$$(1 + \alpha B)(1 + \bar{\alpha} B) = (1 + \nu_1 B + \nu_0 B^2) . \tag{3.12}$$

In terms of the ν's

$$\begin{aligned}(\alpha + B)(\bar{\alpha} + B)H(B) &= \nu_0 H_0 + (\nu_0 H_1 + \nu_1 H_0)B \\ &\quad + (\nu_0 H_2 + \nu_1 H_1 + H_0)B^2 + \cdots\end{aligned} \tag{3.13}$$

and

$$\begin{aligned}(1 + \alpha B)(1 + \bar{\alpha} B)H(B) &= H_0 + (\nu_1 H_0 + H_1)B \\ &\quad + (\nu_0 H_0 + \nu_1 H_1 + H_2)B^2 + \cdots .\end{aligned} \tag{3.14}$$

To obtain the analogs of (3.8) and (3.9), i.e., the left and right sides of (3.11), note that the square of the coefficient of B^j ($j \geqq 2$) above has 6 terms, 3 squares and 3 cross products. Multiplying these out, adding them up, and matching like terms in the expansions of (3.13) and of (3.14), the right side minus the left side of (3.11) is

$$\begin{aligned}(H_{p-1}^2 + H_p^2)(1 - \nu_0^2) + 2H_p H_{p-1}(\nu_1 - \nu_1 \nu_0) \\ = (x^2 + 2\rho xy + y^2)(1 - \nu_0^2)\end{aligned} \tag{3.15}$$

where $x = H_{p-1}$, $y = H_p$, and

$$\rho = \frac{\nu_1 - \nu_1 \nu_0}{1 - \nu_0^2} = \frac{\nu_1}{1 + \nu_0} = \frac{-\phi_2}{1 - \phi_2} \tag{3.16}$$

where $\phi_1 = -\nu_1$, $\phi_2 = -\nu_0$. Recognizing that ρ in (3.16) is the negative of the lag-1 autocorrelation of the second order autoregressive process

$$(1 - \phi_1 B - \phi_2 B^2)y_t = (1 + \alpha B)(1 + \bar{\alpha} B)y_t = e_t , \tag{3.17}$$

which is stationary since $|\alpha| < 1$, it follows that $|\rho| < 1$. Consequently, the quadratic form (3.15) is positive definite, whence nonnegative for all H_{p-1}, H_p. □

To establish the theorem in general, undergo the same sequence as described following (3.10), except that whenever a complex root is encountered, handle it and its conjugate with one application of Lemma 3 rather than two applications of Lemma 2. □

It is also seen from this proof [(3.10) and (3.15)] that $\Delta(p)$ is strictly positive unless H_{p-1} and H_{p-2} are zero for every application of the two lemmas. This will never occur if, for example, $Q(B)$ contains a factor $(1 - \nu B)^{-1}$, $|\nu| < 1$, e.g. if $\delta(B)$ is of degree 1 or if x is first order autoregressive. On the other hand if $Q(B)$ is of finite degree then $\Delta(p) > 0$ for only finitely many p. In any event, it is "short term" forecasts which are most affected, as seen by

THEOREM 4. *For any noninvertible transfer function,*

$$\lim_{p\to\infty} \Delta(p) = 0\,. \tag{3.18}$$

PROOF. The variance of h_t is the same in either its explicit or its implicit representation, this variance being

$$\sigma^2 \textstyle\sum_{j=0}^{\infty} (\chi_j^{(E)})^2 = \sigma^2 \sum_{j=0}^{\infty} (\chi_j^{(I)})^2\,. \tag{3.19}$$

The difference of the two sides of (3.19) is $\lim \Delta(p)$. □

The case of pure delay. Very often in applications there is a delay of b time units before a change in x influences y. Thus, in effect, $\omega(B)$ contains the factor B^b, so that $\alpha_1 = \cdots = \alpha_b = 0$ in (2.1). Thus delay is seen to be but a polar case of noninvertibility in transfer functions. For $p \leqq b$,

$$\Delta(p) = \sigma^2 \textstyle\sum_{j=0}^{p-1} [\chi_j^{(I)}]^2\,, \tag{3.20}$$

which implies that the forecast error in y results only from the error term, which is of course always true when the independent-variable values at the time being forecast are known exactly. The quantity $B^{-b}\omega(B)$ may still be noninvertible; if it is invertible, then in the above treatment, $\nu(B) = B^b$. Two examples where $\nu(B) = B^3$ are described in [1, Chapter 11]; estimated models involving delay plus additional noninvertibility are contained in [4].

REFERENCES

[1] BOX, G. E. P. and JENKINS, G. M. (1970). *Time Series Analysis, Forecasting and Control.* Holden-Day, San Francisco.

[2] PIERCE, D. A. (1972). Least squares estimation in dynamic-disturbance time series models. *Biometrika* **59** 73–78.

[3] WHITTLE, P. (1963). *Prediction and Regulation by Linear Least Squares Methods.* English Universities Press, London.

[4] ZELLNER, A. and PALM, F. (1974). Time series analysis and simultaneous equation models. *J. Econometrics* **2** 17–54.

DIVISION OF RESEARCH AND STATISTICS
FEDERAL RESERVE BOARD
WASHINGTON, D.C. 20551