FEEDBACK BETWEEN STATIONARY STOCHASTIC PROCESSES

by

P. E. Caines and C. W. Chan

Memorandum No. ERL-M460

3 July 1974

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

FEEDBACK BETWEEN STATIONARY STOCHASTIC PROCESSES

P. E. Caines

Department of Electrical Engineering and Computer Sciences
and the Electronics Research Laboratory
University of California, Berkeley, California 94720

C. W. Chan

Control Systems Centre
Control Systems Centre, U. M. I. S. T.
Manchester, England

Abstract

A simple formulation is given for the notion of feedback between two stationary stochastic processes in terms of the canonical representation of the joint process. The definition presented here has consequences in filtering thoery and provides statistical criteria concerning the identification of systems which may contain feedback.

The research of the first author was sponsored by the National Science Foundation, Grant GK-10656X3 and that of the second author was supported by the Science Research Council, Grant B/72/1238.

1. Introduction

In economic, social and biological modelling it is often of interest to inquire whether a feedback mechanism exists between those processes labeled as outputs and those labeled as inputs. There are two principal reasons for this. First, the theory of most system identification techniques, such as least squares or maximum likelihood, require that the observed input processes be statistically independent of the unobserved random processes appearing in the observed output process. Second, it is often of scientific and behavioural interest to know whether some sort of feedback is present in a system. For example, the existence of feedback between money supply and national income has long been debated by econometricians and a similar question arises in the relationship between unemployment and gross domestic product. In the area of biology one may pose the statistical problems of detecting feedback relationships between blood sugar and insulin levels and between stimulus and response processes in the nervous system.

The issues introduced above have long been discussed in the time series literature in terms of the notion of causality. Wiener [1] suggested that one time series should be called causal to a second if knowledge of the first series reduced the mean square prediction error of the second series. Granger and others [2,3] have elaborated upon this approach. According to these authors feedback is said to be present when each of two series is causal to the other. In Section 2 of this paper we introduce an alternative definition of feedback to this one. We should remark here that none of the features of stochastic processes which we isolate will be termed "causality." The reason for this is that we believe the concept of causality belongs properly to the realm of experimental science, while the notions we introduce are meaningful for stochastic processes viewed simply as mathematical objects, and for situations where the experimenter is limited to merely recording selected observations. This is the case, for example, in econometrics and astronomy. Of course, despite our disclaimer, the reader is free to decide in the end that we have merely introduced yet another notion of causality.

It is worth mentioning in the context of this discussion that definitions of causality have been suggested which appear to fit the experimental situation more closely than those of Wiener and Granger. Gersch [4,5], for instance, has proposed a notion of causality between three or more time series that has applications in neurobiology.

The contents of this paper are broadly as follows: in Section 2 we introduce a precise definition of feedback between an ordered pair of multivariate processes in terms of the canonical representation of the joint process with respect to its innovations. This section begins with a discussion which gives strong motivation for our definition of feedback and closes with a list of properties following from that definition. In Section 3 we show there is a relationship between the feedback properties of a process and the causal structure of an associated optimal Wiener filter. Finally, in Section 4, these ideas are applied in system identification. Two examples are presented. In the second two statistical tests support the conclusion that the gross domestic product - unemployment relation is feedback free.

We believe this is the first work to propose and use multivariate testing procedures on the feedback problem. Much of the material presented here first appeared during 1972 in Chan [6]. Independently Sims [7] formulated a definition of causality which also involves the joint input-output process. He presented an argument that his formulation is equivalent to Granger's. Using univariate statistical procedures Sims examines the money-income relationship and concludes that while there is no feedback from income to money supply there is feedback from money to income. Goodhart et al. [8] have recently applied Sims methods to the U.K. economy obtaining the opposite results to those of Sims. Finally Wall [9] has proposed that a specific canonical structure for the joint input-output process should correspond to a unidirectional causal relation between input and output. Wall's criterion turns out to be identical to our own. He concludes from a careful statistical analysis of data for the United Kingdom from 1955 to 1971 that bidirectional causality or, in our terms, feedback exists between the ordered pair of processes money supply and income.

2. Definition and Properties of Feedback

We begin this section by a discussion which will motivate our definition of feedback.

1. Let y and u be p and q component stationary stochastic processes respectively and let $\{K_i, i\geq 0\}$, $\{L_i, i\geq 0\}$, $\{M_i, i\geq 0\}$ and $\{N_i, i\geq 0\}$ be square summable sequences of $q\times p$, $q\times q$, $p\times q$, $p\times p$ matrices respectively. Further let v and w be respectively p and q component, zero mean, independent identically distributed (i.i.d.) processes which are possibly correlated. Then the following equations are often taken to represent the dynamical behaviour of a feedback system:

$$y_t = \sum_{i=0}^{\infty} K_i u_{t-i} + \sum_{i=0}^{\infty} L_i v_{t-i} \tag{2.1a}$$

$$u_t = \sum_{i=0}^{\infty} M_i y_{t-i} + \sum_{i=0}^{\infty} N_i w_{t-i} \tag{2.1b}$$

Equation (2.1a) gives the relationship between the observed input process u, the unobserved process v, which is generally known as the output disturbance, and the output of the dynamical system y. Equation (2.1b) is interpreted as describing the behaviour of a feedback loop since it gives the relationship between the output y, unobserved disturbance w and the input process u. Akaike [10] and Bohlin [11] define the absence of feedback for a system of the form (2.1a) as the independence of the processes u and v. This is reasonable since intuition associates a lack of feedback from the observed process y to the observed process u with the notion that u is in some sense exogenous or external to the process generating y i.e. y is a function of the process u and an an unobserved process v with the σ-fields generated by u and v being independent.

The problem with using (2.1a) alone to define the lack of feedback between y and u is that the criteria usually proposed depend upon v which is unobserved and cannot be estimated without a-priori knowledge of the parameters of the system. On the other hand the formulation we introduce below is consistent with the ideas of Akaike and Bohlin but is formulated in terms of observable processes.

Let $\underline{H}$ denote the Hilbert space which is the mean square completion of the space of all processes with finite first and second moments. Further let K etc. denote the map from $\underline{H}$ into $\underline{H}$ given by $Kv = \{\sum_{i=0}^{\infty} K_i v_{t-i};\ -\infty < t < \infty\}$. Then, formally solving (2.1) for y and u in terms of the processes v and w, we obtain

$$\begin{bmatrix} y \\ u \end{bmatrix} = \begin{bmatrix} (I-KM)^{-1}L & (I-KM)^{-1}KN \\ (I-MK)^{-1}ML & (I-MK)^{-1}N \end{bmatrix} \begin{bmatrix} v \\ w \end{bmatrix}$$

$$= \begin{bmatrix} PL & PKN \\ QML & QN \end{bmatrix} \begin{bmatrix} v \\ w \end{bmatrix}, \tag{2.2}$$

where KM etc. denotes the concatenation of the operators K and M and $P \underline{\Delta} (I-KM)^{-1}$ and $Q \underline{\Delta} (I-MK)^{-1}$, where the inverse operators are assumed to exist.

Let us assume that feedback is absent from (2.1) in the sense that M is identically zero. Then (2.2) reduces to

$$\begin{bmatrix} y \\ u \end{bmatrix} = \begin{bmatrix} L & KN \\ 0 & N \end{bmatrix} \begin{bmatrix} v \\ w \end{bmatrix}$$

and y has the representation

$$y = Ku + Lv$$

in terms of the independent processes u and v which are respectively observable and unobservable. This discussion suggests that the triangular structure of the joint representation (2.2) can be used to formulate the notion of the absence of feedback. However equation (2.2) was derived from equation (2.1), and, as it stands, this is not a unique (canonical) representation. For this reason we begin the discussion afresh starting from the theory of stationary stochastic processes.

2. Consider the r component process ζ lying in $\underline{H}$. (For convenience in this paper we shall take all stochastic processes to have zero mean.) Let $\underline{H}_m^n$ denote the subspace of $\underline{H}$ generated by $\{\zeta_m,\ldots,\zeta_n\}$ and let $\eta|\underline{H}_m^n$ denote the projection of a random variable η on H_m^n. When $\lim_{n\to-\infty} \zeta_t|\underline{H}^n_{-\infty} = 0$ the process ζ has zero projection on the infinite past and is termed regular [12] or non-deterministic [13]. Further, when the process ζ has the property that $E\zeta_n\zeta_n^T = \Sigma_n > 0$, the process ζ is said to be of full rank. It was shown by Wold that a regular full rank stationary process possesses a canonical one sided moving average representation of the form

$$\zeta_t = \Phi_0\, \varepsilon_t + \Phi_1\, \varepsilon_{t-1} + \ldots.. \tag{2.3}$$

where ε is an r component stationary orthonormal process (i.e. $E\varepsilon_t\varepsilon_s' = I\delta_{ts}$), Φ_0 is invertible and the subspaces spanned by $\{\zeta_t^i;\ t\leq n,\ i=1,\ldots,r\}$ and $\{\varepsilon_t^i;\ t\geq n,\ i=1,\ldots,r\}$ are identical. In order to make the representation

(2.3) unique we take ϕ_0 to be upper triangular with positive elements on the diagonal. (Otherwise the sequence $\{\Phi_0,\Phi_1,...\}$, together with the orthonormal process $U\varepsilon$, and the sequence $\{\Phi_0 U,\Phi_1 U,...\}$, together with the orthonormal process ε, yield two different canonical representations of the same process, when $UU^T = I$.)

In this paper we shall only deal with finitely generated stationary processes; these constitute the subclass of regular full rank stationary processes for which there exist two sets of real matrices $\Delta = \{\Delta_1,...,\Delta_n\}$ and $\Gamma = \{\Gamma_0,\Gamma_1,...,\Gamma_n\}$ such that

$$\zeta_t + \Delta_1 \zeta_{t-i} + \dots + \Delta_n \zeta_{t-n} = \Gamma_0 \varepsilon_t + \dots + \Gamma_n \varepsilon_{t-n} \tag{2.4}$$

From this point on the phrase 'stationary stochastic process' shall denote only a regular full rank finitely generated stationary stochastic process. Let

$$\Delta(z) = I + \Delta_1 z^{-1} + \dots + \Delta_n z^{-n}$$

and

$$\Gamma(z) = \Gamma_0 + \Gamma_1 z^{-1} + \dots + \Gamma_n z^{-n}$$

where $z \in \mathbb{C}$ the field of complex numbers. Clearly $\Delta^{-1}(z)$ exists at all except a finite number of values of the argument. In order to present the first definition we use the following block decomposition of the matrix of rational functions $\Delta^{-1}(z)\ \Gamma(z)$; we shall write

$$\begin{bmatrix} A(z) & B(z) \\ C(z) & D(z) \end{bmatrix}$$

where $A(z)$, $B(z)$, $C(z)$, $D(z)$ are matrices of rational functions of dimension $p\times p$, $p\times q$, $q\times p$, $q\times q$ respectively. Then, motivated by the discussion in Section 2.1, we make the following

<u>Definition 1</u>. Let ζ be a p+q component stationary stochastic process composed of the p component process y and the q component process u. Further let $\Phi(z) = \Delta^{-1}(z)\ \Gamma(z) = \Phi_0 + \Phi_1 z^{-1} + \Phi_2 z^{-2} + \ldots$. Then we say there is no feedback from y to u or the process ζ is (p,q) feedback free if and only if the matrix $\Phi(z)$ has the form

$$\begin{bmatrix} A(z) & B(z) \\ 0 & D(z) \end{bmatrix} \tag{2.5}$$

If ζ is not (p,q) feedback free we say there is (p,q) feedback from y to u.

It is convenient to introduce immediately an equivalent definition to the one given above. This second definition will prove to be more useful than the first in the context of system identification.

Consider the process ε in the canonical representation (2.3). Construct the r component stationary i.i.d. process δ by setting $\delta_t = \Phi_0 \varepsilon_t$. Then ζ has the canonical representation

$$\zeta_t = \zeta_t + \Theta_1\ \zeta_{t-1} + \Theta_2\ \zeta_{t-2} + \ldots \tag{2.6}$$

where $E\ \delta_t \delta_t' = \Phi_0 \Phi_0'$. Notice that $\Theta(\infty) = I$ when $\Theta(z) \underline{\Delta} \sum_{i=0}^{\infty} \Theta_i z^{-i}$ with $\Theta_0 \underline{\Delta} I$. Using (2.6) we state

<u>Definition 2</u>. The process ζ is (p,q) feedback free if and only if the matrix $\Theta(z)$ derived from the representation (2.6) is upper block triangular with indices {p,q}.

Since $\Theta(z)\Phi_0 = \Phi(z)$ and since Φ_0 is upper triangular, has positive elements on the diagonal and is invertible Definition 1 and 2 are clearly equivalent.

Zero mean stationary gaussian processes are completely characterized by their spectral density matrices $\Psi(z)$. For such processes the property of being feedback free is given by the structure of the spectral density matrix:

<u>Proposition 1</u>

The p+q component stochastic process ζ is (p,q) feedback free if and only if there exist matrices of rational functions $A(z)$, $B(z)$, $D(z)$ of dimension $p\times p$, $p\times q$, $q\times q$ respectively with $A(z)$, $D(z)$ of full, normal rank (i.e. full rank except at a finite number of points) and the poles of $A(z)$, $B(z)$, $D(z)$, $A^{-1}(z)$ and $D^{-1}(z)$ lying in the open unit disc, such that

$$\Psi(z) = \begin{bmatrix} A(z)\ A^{*}(z) + B(z)\ B^{*}(z) & B(z)\ D^{*}(z) \\ D(z)\ B^{*}(z) & D(z)\ D^{*}(z) \end{bmatrix}, \tag{2.7}$$

where $A^{*}(z)$ denotes $A'(z^{-1})$ etc.

Let us describe a matrix of rational functions as (asymptotically) stable when all its poles lie in the open unit disc. Then a theorem due to Youla [14] states that the spectral density matrices of full rank stationary processes possess stable and inverse stable spectral factors. The proposition follows immediately using this result.

A causal operator T shall be called a rational operator on $\underline{H}$ when $T\zeta = \{\sum_{i=0}^{\infty} T_i \zeta_{t-i};\ -\infty < t < \infty\}$ and $T(z) = \sum_{i=0}^{\infty} T_i z^{-i}$ is rational and all poles of $T(z)$ lie in the open unit disc. Employing this terminology, the reader may readily verify the first four of the following simple properties of our notion of feedback:

Some Properties of Feedback

1. Let ζ be a (p,q) feedback free process. Then operations with any upper block triangular rational operator T(i.e. $T(z)$ upper block triangular with indices $\{r,s\} \times \{p,q\}$) maps ζ into an (r,s) feedback free process $T\zeta$ when the blocks on the diagonal of $T(z)$ are of full normal rank. (The last condition ensures the image process is full rank. It can be dropped if we relax the full rank assumption.)

2. In general operations with matrices of rational operators introduce feedback into a feedback free process. However given a process ζ with (p,q) feedback there exists a lower block triangular rational operator T (indices $\{p,q\}$) such that $T\zeta$ is feedback free.

3. Let w be a p component zero mean stationary process (not necessarily of full rank) which is independent of the process $[y',u']$. If the ordered pair of processes $\langle y,u\rangle$ is feedback free (i.e. $\zeta' = [y',u']$ is feedback free) then so is $\langle y + w,u\rangle$. However, if w is a q component process it is in general not the case that $\langle y,u + w\rangle$ is feedback free.

4. $\langle y,u\rangle$ and $\langle u,y\rangle$ are feedback free, and y_t is independent of u_t for all t, if and only if the process y is independent of the process u.*

5. $\langle y,u\rangle$ is feedback free if and only if there exists a unique representation of y in the form

$$y = Ku + Lv \tag{2.8}$$

where K and L are rational operators with L invertible, L_0 is upper triangular with positive elements on the diagonal, the processes u and v are

*We point out that the notion of feedback only becomes significant when a system contains dynamics e.g. $y_t=\varepsilon_t^1 + \varepsilon_{t-1}^2$, $u_t=\varepsilon_t^2$ is feedback free when ε^1, ε^2 are independant i.i.d. processes. In the 'static' case, $[y',u']$ an i.i.d. Gaussian process with covariance Σ, $\langle y,u\rangle$ and $\langle u,y\rangle$ are always feedback free.

independent and v is an orthonormal process.

This is shown as follows. From Definition 1 $\langle y,u \rangle$ is feedback free if and only if

$$\begin{bmatrix} y \\ u \end{bmatrix} = \begin{bmatrix} L & M \\ 0 & N \end{bmatrix} \begin{bmatrix} v \\ w \end{bmatrix}$$

with $[v',w']$ an orthonormal process. Now a property of the unique representation (2.3) is that the rational operator Φ is invertible and hence that Φ has a causal stable inverse. (In the Gaussian case this is equivalent to saying that the innovations representation (2.3) gives the stable minimum phase factor of the spectral density matrix $\Phi(z)$).

As a consequence

$$y = MN^{-1}u + Lv,$$

where MN^{-1} and L are rational operators, L is invertible, L_0 has the required form and u and v are independent with v orthonormal.*

Clearly if y has such a representation then $\langle y,u \rangle$ is feedback free.

Uniqueness of the representation (2.8) is demonstrated quite simply. Assume

$$Ku + Lv = \tilde{K}u + \tilde{L}\tilde{v}.$$

Then

$$(K-\tilde{K})u = \tilde{L}\tilde{v} - Lv \tag{2.9}$$

*The reader should notice that for any pair of correlated processes a representation of the form (2.8) exists with the condition that v_t is independent only of u_t. This representation is obtained by setting $[Ku]_t = \hat{y}_{t|t-1}$, $[v]_t \stackrel{\Delta}{=} y_t - \hat{y}_{t|t-1}$ and L = I. However the entire process v is not independent of the entire process u.

Since u is independent of v and $\tilde{v}$ and all processes are zero mean (2.9) yields $K = \tilde{K}$. Now consider the process $Lv = \tilde{L}\tilde{v}$. By assumption L and $\tilde{L}$ are invertible rational operators with L_0 and $\tilde{L}_0$ upper triangular with positive elements on the diagonal and v and $\tilde{v}$ are orthonormal processes. It follows that Lv and $\tilde{L}\tilde{v}$ are both innovations representative of the same process and hence $L = \tilde{L}$ and $v = \tilde{v}$.

We shall mention in passing two other extensions of the concept of feedback free systems. We do not pursue these generalizations in this paper because of their limited scope for applications. First, we may define a non-stationary process ζ to be asymptotically (p,q) feedback free if it is generated by

$$\zeta_t + \Delta_{1,t}\zeta_{t-1} + \dots + \Delta_{n,t}\zeta_{t-n} = \Gamma_{0,t}\varepsilon_t + \dots + \Gamma_{n,t}\varepsilon_{t-n}$$

and

$$\lim_{t\to\infty} [I + \Delta_{1,t}z^{-1} + \dots + \Delta_{n,t}z^{-n}]^{-1} [\Gamma_{0,t} + \dots + \Gamma_{n,t}z^{-n}]^{-1} = \Delta^{-1}(z)\Gamma(z)$$

where $\Delta^{-1}(z)\Gamma(z)$ has the structure (2.5). Second, a family of processes given in a preassigned order $\{y_1,\dots,y_k\}$ might be defined as feedback free if the canonical representation of the vector process $[y_1',\dots,y_k']$ has a canonical representation possessing an upper block triangular structure with indices $\{\dim y_1,\dots,\dim y_k\}$.

Finally we remark that the definitions and results above are also applicable to continuous time processes. The causal rational operators of this section are merely replaced with stationary causal stochastic integrals $\int_0^\infty h(t-s)\, dw_s$ where w is a Wiener process and the kernel function $\{h(\tau), 0\le\tau<\infty\}$ has a rational Fourier transform. The matrices

(2.5), (2.6) and (2.7) appearing in the definitions are unaltered except in their arguments and the definition of conjugation.

3. Linear Least Squares Estimation and Feedback

Given any two processes y and u for which the autocorrelation function $R_u(i,j)$ and the cross correlation function $R_{yu}(i,j)$ are defined one may formulate the corresponding Wiener filtering problem i.e. find the impulse response $\{h(i,j);\ i,j = \text{integer}\}$ of the linear system which when driven by the process u has output $\hat{y}$ which minimizes Trace $E(y(t)-\hat{y}(t))(y(t)-\hat{y}(t))'$ for each integer t. It is well known [15-17] that this leads to the Wiener-Hopf equation

$$\sum_{r=-\infty}^{\infty} h(t,r)\, R_u(r,s) = R_{uy}(t,s) \tag{3.1}$$

In the case of non-stationary autoregressive processes the causal solution to (3.1) may be given in terms of the spectral factors of infinite matrices [18]. In the stationary case, without the restriction that $\{h(t);\ t = \text{integer}\}$ be causal, (3.1) has the solution

$$H(z) = \Psi_{yu}(z)\Psi_u^{-1}(z) \tag{3.2}$$

where $\Psi_{yu}(z)$ and $\Psi_u(z)$ are the z-transforms of $R_{yu}(t)$ and $R_u(t)$ and H(z) is the z-transform of h(t).

To solve the problem when the filter is restricted to be causal we use the following assumptions and notation. Each discrete transform A(z) will be assumed analytic in some annulus $\rho < \|z\| < \rho^{-1}$ $(0 < \rho < 1)$ containing the unit circle. (In particular, this implies that all spectral densities are analytic on the unit circle.) Furthermore the causal truncation of the Laurent expansion of any A(z) will be assumed to be analytic in the region $1 \le |z|$ and the anticausal truncation analytic in $|z| \le 1$ i.e. when

$$A(z) = \sum_{i=-\infty}^{\infty} A_i z^{-i} \quad , \quad \rho < \|z\| < \rho^{-1},$$

then

$$[A(z)]_+ = \sum_{i=0}^{\infty} A_i z^{-i}$$

converges for $1 \leq \|z\|$, and

$$[A(z)]_- = \sum_{i=-\infty}^{-1} A_i z^{-i}$$

converges for $\|z\| \leq 1$. This is equivalent to saying that A(z) only represents causal sequences that are summable (and hence stable in the sense that $\|A_i\| \to 0$ as $i \to \infty$) and anti-causal sequences that are summable (and hence anti-stable in the sense that $\|A_i\| \to 0$ as $i \to -\infty$). These assmuptions allow us to identify the operation of causal truncation of any rational A(z) with the operation of selecting that part of the partial fraction expansion of A(z) which has all its poles in the open unit disc (see [19]). We remark that this association of causality and stability by identifying the contour of inversion with the boundary of the stability region is a convenient simplifying assumption often employed in continuous time problems (see e.g. [15]).

Using the notation just introduced the optimal causal Wiener filter has an impulse response given by

$$H_+(z) = [\Psi_{yu}(z)\ (\Theta_u^*(z))^{-1}]_+\ \Theta_u^{-1}(z) \tag{3.3}$$

where $\Theta_u(z)$ is a stable and an inverse stable spectral factor of $\Psi_u(z)$

and $\Theta_u^*(z)$ denotes $\theta_u^T(z^{-1})$.

It is interesting to ask whether the structure of the optimal causal Wiener filter for estimation of the process y from the process u is affected by the feedback properties of the joint process (y,u). In answer to this question we obtain the following result.

<u>Theorem</u> If the processes (y,u) are feedback free the optimal causal Wiener filter for the estimation of y from u is identical to the optimal non-causal filter. Conversely assume the optimal causal and non-causal Wiener filters for the estimation of y from u are identical and assume

(A1) The poles of B(z), C(z), D(z) in (2.7) and the stable zeros of $|A(z)A^*(z)+B(z)B^*(z)|$ (i.e. the zeros of $|\theta(z)|$) form pairwise disjoint sets.

(A2) The poles of C(z) each have multiplicity one and at the poles of $C^*(z)$ the matrices A(z) and $\hat{\theta}^*(z)$ have full rank.

Then $\langle y,u \rangle$ is feedback free.

<u>Proof</u>. When $\langle y,u \rangle$ is feedback free express $\Psi_{yu}(z)$ and $\Psi_u(z)$ in terms of the matrices of the canonical representation (2.3). Then

$$\Psi_{yu}(z) = B(z)\, D^*(z),$$

$$\Psi_u(z) = D(z)\, D^*(z).$$

This yields immediately

$$H_+(z) = [B(z)\ D^*(z)\ (D^*(z))^{-1}]_+\ D^{-1}(z)$$

$$= B(z)\ D^{-1}(z)$$

$$= H(z)$$

which proves sufficiency.

Necessity is only slightly less simple. Assume the causal and non-causal filters are equal i.e. $[\Psi_{yu}(z)(\Theta_u^*(z))^{-1}]_+\ \Theta_u^{-1}(z) = [\Psi_{yu}(z)(\Theta_u^*(z))^{-1}]$ $\Theta_u^{-1}(z)$

then

$$[\Psi_{yu}(z)\ (\Theta_u^*(z))^{-1}]_+ = \Psi_{yu}(z)\ (\Theta_u^*(z))^{-1}$$

or, equivalently

$$[\Psi_{yu}(z)\ (\Theta_u^*(z))^{-1}]_- = 0$$

Let $\hat{\Theta}_u(z)$ denote $\Theta_u^{-1}(z)$ and $\hat{\Theta}_u^*(z)$ denote $(\Theta_u^*(z))^{-1}$. Further for rational matrices M(z), N(z) let R_M($R_{M,N}$ respectively) denote that part of the partial fraction expansion of $\Psi_{yu}(z)\ \hat{\Theta}^*(z)$ which has poles in common with M(z) (M(z) or N(z) respectively). Since A(z), B(z), C(z), D(z) and $\Theta(z)$ are causal it follows that

$$\begin{aligned}
0 &= [\Psi_{yu}(z)\ \hat{\Theta}^*(z)]_- \\
&= [(A(z)\ C^*(z) + B(z)\ D^*(z))\ \hat{\Theta}^*(z)]_- \\
&= [R_{A,B}(z) + R_{C*}(z) + R_{D*}(z) + R_{\hat{\Theta}*}(z)]_- \\
&= R_{C*}(z) + R_{D*}(z) + R_{\hat{\Theta}*}(z).
\end{aligned}$$

Using (A1) we conclude that $R_{C*}(z) = R_{D*}(z) = R_{*}(z) = 0$. Now $A(z)$, $B(z)$ and $D(z)$ can have no poles in common with $C^{*}(z)$ and by (A1) the same is also true of $\hat{\Theta}^{*}(z)$ and $C^{*}(z)$. For $R_{C*}(z)$ this yields

$$R_{C*}(z) = \sum_{i=1}^{c} \frac{1}{(z-\nu_i)} \lim_{z\to\nu_i} (z-\nu_i) \; [(A(z)C^{*}(z)+B(z)D^{*}(z))\hat{\Theta}^{*}(z)]$$

$$= \sum_{i=1}^{c} \frac{1}{(z-\nu_i)} \; A(\nu_i)C^{*}_{\nu_i} \; \hat{\Theta}^{*}(\nu_i)$$

where $\{\nu_i;\ 1\leq i\leq c\}$ are the set of poles of $C^{*}(z)$, $C^{*}_{\nu_i}$ denotes the residue of $C^{*}(z)$ at the pole ν_i and where we have used (A1) again. Since by (A2) $A(z)$ and $\hat{\Theta}^{*}(z)$ are full rank at the poles of $C^{*}(z)$ the residue matrices of $C^{*}(z)$ vanish at each of its poles. Consequently $C(z) = 0$ and the result follows.

Incidentally when $C(z) = 0$ we have $\hat{\Theta}^{*}(z) = (D^{*}(z))^{-1}$ and so $\Psi_{yu}(z)\ \hat{\Theta}^{*}(z) = B(z)D^{*}(z)(D^{*}(z))^{-1} = B(z)$. But then clearly by (A1) $R_{C*}(z) = R_{D*}(z) = R_{\hat{\Theta}*}(z) = 0$.

Because the spectrum of a feedback free process has a special structure certain standard formulae associated with stationary processes take on a very simple form in the feedback free case. Two examples are the following:

1. Let the processes $\langle y,u \rangle$ be feedback free, then, as above, $\Psi_y(z) = A(z)A^{*}(z) + B(z)B^{*}(z)$ and $\Psi_u(z) = D(z)D^{*}(z)$. Substituting these expressions into the standard formula [10] for the mean square filtering error yields:

$$E\|y_t-\hat{y}_{t|t}\|^2 = \text{Trace}\ \frac{1}{2\pi j}\int_{|z|=1} (\Psi_y(z) - H(z)\Psi_u(z)H^*(z))\ z^{-1}\,dz$$

$$= \text{Trace}\ \frac{1}{2\pi j}\int_{|z|=1} A(z)A^*(z)\ z^{-1}\,dz \tag{3.4}$$

(3.3) and (3.4) show that in the feedback free case the optimal causal (equivalently non-causal) filter is independent of A(z) and the mean square filtering error is independent of B(z) and D(z).

2. The mutual information between the processes y and u is given by the following expression [20]:

$$J(y,u) = \frac{1}{2\pi j}\int_{|z|=1} \log \frac{|\Psi_y(z)|}{|(\Psi_y(z) - \Psi_{yu}(z)\Psi_u^{-1}(z)\Psi_{yu}^*(z))|}\ \frac{dz}{z}$$

When $\langle y,u \rangle$ is feedback free this formula also simplifies yielding:

$$J(y,u) = \frac{1}{2\pi j}\int_{|z|=1} \log \frac{\det|A(z)A^*(z)+B(z)B^*(z)|}{|A(z)A^*(z)|}\ \frac{dz}{z}$$

It can be seen that unlike the case where feedback is present the expression for the mutual information between output and input does not depend on D(z).

4. Detection of Feedback

As we described in the Introduction there are at least two reasons for wishing to detect feedback in a system whose parameters are being estimated. First, there is the simple fact that the existence of feedback may be of scientific and engineering interest. Second, most recent analyses [21-22] of the properties of parameter estimates depend on the assumption that the observed input process is independent of the system disturbances, in other words it assumed the system is feedback free. When this is not the case it may well happen that estimates, such as the maximum likelihood estimate, are not consistent and are biased. Recently Ljung [24] has given conditions for prediction error estimation methods to generate parameter estimates that fall into a particular equivalence class of the true parameters. However this equivaluence class is, in its turn, defined in terms of a prediction error criterion.

Models of the form

$$y = Ku + Lv \tag{4.1}$$

$$u = My + Nw \tag{4.2}$$

are frequently postulated for the identification of closed loop systems (see e.g. [25]). If it is assumed that the unobserved processes v and w are independent then it is possible to make (4.1), (4.2) unique. However when (4.1),(4.2) is used in an identification experiment, or when (4.1) or (4.2) are used separately, it is not possible to impose this independence assumption. Without this assumption the representation (4.1), and hence the entire model (4.1)-(4.2), suffers from an inherent non-uniqueness. This may be shown as follows. Let $\tilde{K}$ be a rational operator.

Then using (4.1) and (4.2) we may write

$$y = (K-\tilde{K})\ u + \tilde{K}u + Lv$$
$$= (K-\tilde{K})\ u + \tilde{K}\ My + KNw + Lv$$

For any such $\tilde{K}$ this yields a representation of y in the form

$$y = (I-\tilde{K}M)^{-1}\ (K-\tilde{K})\ u + x$$

where $(I-\tilde{K}M)$ is assumed inverse causal and stable and the spectral density of x is

$$\Psi_x = (I-\tilde{K}M)^{-1}\ [\tilde{K}NN^*\tilde{K}^* + LL^*]\ (I-\tilde{K}M)^{*-1}$$

Clearly there exists an entire family of operators $\{\tilde{K}\}$ that give rise to a family of representations of y in the form (4.1). In addition to models of the form (4.1), (4.2), which contain unobserved noise in the equation for the input u, models with deterministic feedback are frequently postulated in the literature. Such models are obtained by setting $N = 0$ in equation (4.2). In this case the model (4.1), (4.2) is invariably non-unique. This may be seen by merely setting $N = 0$ in $\Psi_x(z)$ above. As has been suggested in the literature, uniqueness may be retrieved for (4.1) by placing conditions on the orders of the rational operators K and L. Unfortunately it is precisely this information which is unknown at the commencement of an identification exercise.

Other non-uniqueness properties of models of the form (4.1), (4.2) have been discussed by Akaike [26]. In this paper by using the unique innovations representation for the joint process we dispose of the problem of ambiguous system representations. The detection of feedback in practical situations then reduces to the application of hypothesis testing techniques to the estimated structure of the system generating the joint input-output process. Once it is established that the joint process is feedback free Property 5 of Section 2 assures us that we may consistently estimate the parameters of a unique representation of the form (2.8).

4.1. System Identification

Let the Gaussian process ζ be generated by the model (2.4) which we write again here for convenience:

$$\zeta_t + \Delta_1\zeta_{t-1} + \dots + \Delta_n\zeta_{t-n} = \Gamma_0\varepsilon_t + \dots + \Gamma_n\varepsilon_{t-n},$$

where ε_t is a Gaussian orthonormal process. Since ζ is also Gaussian so is the innovations process $\{e_t\} = \{\zeta_t - \hat{\zeta}_{t|t-1}\}$ of ζ. Using Baye's rule and the independence of the innovations process we obtain

$$P_\theta(\zeta^N) = \prod_{t=1}^{N} P\,(\zeta_t|\zeta^{t-1})$$

$$= \prod_{t=1}^{N} P_\theta(\hat{\zeta}_{t|t-1} + e_t|\zeta^{t-1})$$

$$= \prod_{t=1}^{N} \frac{1}{(2\pi)^{p/2}} \frac{1}{|\Sigma_t|^{1/2}} \exp - \frac{1}{2}\,(e_t^T\,\Sigma_t^{-1}\,e_t).$$

where the filtering error e_t is distributed $N(0,\Sigma_t)$ for $t \geq 1$ [27]. It follows that, up to a constant, the log-likelihood function is given by

$$L^N(\zeta^N,\theta) = -\frac{1}{2}\sum_{t=1}^{N} \log|\Sigma_t| - \frac{1}{2}\sum_{t=1}^{N} e_t^T\,\Sigma_t^{-1} e_t. \tag{4.4}$$

The maximum likelihood estimate $\hat{\theta}^n(\zeta^N)$ of θ is generated by maximizing (4.6) with respect to θ. It is well known [28,29] that maximum likelihood estimates (MLEs) have the desirable properties of strong consistency, asymptotic efficiency and asymptotic normality for the case of independent observations. Dynamical systems clearly possess dependent output processes and so an extension of the classical theory and techniques is required. This was first carried out for scalar autoregressive moving average models by Åström and Bohlin [21]. Åström and Bohlin used the law of large numbers in conjunction with the technique of Kendall and Stuart [30] to prove the strong consistency of the sequence of maximum likelihood estimates $\{\hat{\theta}^n, n=1,2,\ldots\}$. Their proof was incomplete but Rissanen and Caines [31,32] used their suggestion of employing the ergodic theorem to produce another demonstration of the strong consistency of the estimates. It turns out that without knowledge of the Kronecker indices [33] of the rational matrix $\mathring{\Delta}^{-1}(z)\, \mathring{\Gamma}(z)$ neither the parameters of $(\mathring{\Delta},\mathring{\Gamma}) = (\mathring{\Delta}_1,\ldots,\mathring{\Delta}_n,\mathring{\Gamma}_0,\ldots,\mathring{\Gamma}_n)$ nor the parameters of a Markovian state space representation may be consistently estimated, see [31,32]. (We have used here the zero superscript to denote parameter values in the same equivalence class as the true parameters.) However without this knowledge the following result still holds.

<u>Theorem 1</u> [31,32]

Let S be a closed and bounded subset of the space of parameters $\Theta = \{\theta = (\Delta,\Gamma) | \theta \in \mathbb{R}^{(2n+1)r^2}\}$ such that for all $\theta \in S$ the roots of $|\Delta(z)|$ lie in the open unit disc and such that $\Gamma_0(\theta)$ is non-singular and upper trinagular with positive elements on the diagonal. Suppose $\hat{\theta}^N = (\hat{\Delta}^N,\hat{\Gamma}^N)$ maximizes $L_N(\zeta^N,\Theta)$ over S, then

$$\sum_{i=0}^{\infty} \|\hat{\Phi}_i^N - \mathring{\Phi}_i\| \to 0 \text{ a.s. as } N \to \infty$$

where $\Phi(z) = \sum_{i=0}^{\infty} \mathring{\Phi}_i z^i = \mathring{\Delta}^{-1}(z)\mathring{\Gamma}(z)$ and where $\|\cdot\|$ denotes any matrix norm.

In other words the maximum likelihood method produces strongly consistent estimates of the true impulse response. Furthermore, by elaborating Åström and Bohlin's [21] extension of the classical result [29], we have

Theorem 2

Let the conditions of Theorem 1 hold for the Gaussian process ζ, let the Kronecker indices of $\mathring{\Delta}^{-1}(z)\ \mathring{\Gamma}(z)$ be known and let $(\Delta(z),\Gamma(z))$ be left relatively prime for $\theta \in S$. Then the parameter estimate $\hat{\theta}^N$ is strongly consistent at $\hat{\theta}$. Further let $L(\theta)$ denote $\lim_{N\to\infty} \frac{1}{N} L^N(\zeta^N,\theta)$, assume $L(\theta)$ is twice differentiable with respect to θ and assume that $L_{\theta\theta}(\theta)$ is invertible. Then $\hat{\theta}^N$ is asymptotically normal $N(\mathring{\theta},-L_{\theta\theta}^{-1}(\mathring{\theta}))$ where $L_{\theta\theta}(\theta) = \lim_{N\to\infty} \frac{1}{N} L_{\theta\theta}^N(\zeta^N,\theta) = \lim_{N\to\infty} \frac{1}{N} EL^N(\zeta^N,\theta)$ with probability 1.

In practical numerical work the expression (4.4) is approximated by

$$L^N(\zeta^N,\theta) = \frac{-N}{2} \log |\Sigma| - \frac{1}{2} \sum_{t=1}^{N} \tilde{e}_t^T \Sigma^{-1} \tilde{e}_t \tag{4.5}$$

where $\tilde{e}_t$ is computed using the predictor for $\hat{\zeta}_{t|t-1}$ in its steady state form and Σ is the corresponding steady state error covariance matrix. Given a sample $\{\zeta_1 \ldots \zeta_N\}$ the numerical maximization of (4.5) may be carried out in several ways. The method we adopt is the following: first, maximize (4.5) with respect to Σ yielding

$$\hat{\Sigma}(\theta) = \frac{1}{N} \sum_{t=1}^{N} \tilde{e}_t(\theta)\, \tilde{\varepsilon}_t^T(\theta)$$

and

$$L^N(\zeta^N, \hat{e}(\theta), \theta) = -\frac{N}{2} \log |\hat{\Sigma}(\theta)| + \text{const.}.$$

Second, minimize the function

$$V(\theta) = |\hat{\Sigma}(\theta)| \tag{4.6}$$

with respect to θ yielding $\hat{\theta}^N(\zeta^N)$.

4.2 Hypothesis Testing

Using the estimation techniques described above the problem of detecting feedback is reduced to that of choosing between the following models,

$$S_0: \quad Z_0(z) = \Delta_0^{-1}(z)\, \Gamma_0(z) = \begin{bmatrix} A(z) & B(z) \\ 0 & D(z) \end{bmatrix}, \quad Z_0(\infty) = I_{p+q}$$

$$S_1: \quad Z_c(z) = \Delta_c^{-1}(z)\Gamma_c(z) = \begin{bmatrix} A(z) & B(z) \\ C(z) & D(z) \end{bmatrix}, \quad Z_c(\infty) = I_{p+q}$$

where the subscript 0 denotes open loop system (i.e. feedback free), the subscript c denotes closed loop system (i.e. feedback present) and $Z(\infty) = I_{p+q}$ in both cases since we shall use the canonical representation (2.6). Given a system which is feedback free the estimate of $C(z)$ based on input-output records of finite length will naturally not be identically zero. Consequently it is necessary to develop suitable hypothesis testing methods.

Likelihood Ratio Test

Let θ_0 and θ_c be the parameter vectors of S_0 and S_1 and let these vectors have n_0 and n_c components respectively. Then we must choose between the following alternative hypotheses:

$$\left.\begin{aligned} H_0: &\quad \theta = \theta_0 \\ H_1: &\quad \theta = \theta_c \end{aligned}\right\} \tag{4.7}$$

Let ζ^N denote $\{(y_1,u_1),(y_2,u_2),\ldots,(y_N,u_N)\}$ and let $f(\zeta^N,\theta_0)$ and $f(\zeta^N,\theta_c)$ denote the probability density functions of ζ^N with respect to the measures induced by $\theta = \theta_0$ and $\theta = \theta_c$. Then the likelihood ratio is defined by

$$\lambda = \frac{f(\zeta^N,\theta_c)}{f(\zeta^N,\theta_0)} \tag{4.8}$$

and the likelihood ratio test is performed upon the computed value of (4.8).

In the case of independent observations it is known [29] that λ can also be expressed as

$$\lambda = \left(\frac{V(\theta_c)}{V(\theta_0)}\right)^{N/2} \tag{4.9}$$

where $V(\theta)$ is given by (4.6). Further (6.9) can be rewirtten as

$$\lambda = \left(1 + \left(\frac{n_c - n_o}{N - n_c}\right) t\right)^{-N/2}$$

where

$$t = \frac{(N-n_c)}{(n_c-n_0)} \frac{V(\theta_0)-V(\theta_c)}{V(\theta_c)} \tag{4.10}$$

where t is an F-distributed statistic with $(N-n_c, n_c-n_0)$ degrees of freedom i.e. $t \sim F(N-n_c, n_c-n_0)$. Since λ is a monotone decreasing function of t we can translate the likelihood ratio test on λ into an equivalent F-test on t. Applying these results to the dependent sample case of our problem we formulate the following decision rule:

Accept H_0 (Reject H_1) if $t \leq F(N-n_c, n_c-n_0 | H_0)$

Accept H_1 (Reject H_0) if $t > F(N-n_c, n_c-n_0 | H_0)$

where

$$\text{Prob } (t > F(N-n_c, n_c-n_0)\ H_0)) = \alpha$$

and α is the prespecified level of risk.

A similar technique to this has been proposed by Åström [34].

Since the observed processes in our problem are generated by an autoregressive moving average model the validity of this test for the detection of feedback is uncertain. However lacking alternative methods we proceeded in order to see if the F-test would give results consistent with the confidence interval technique described below.

<u>Chi-squared (χ^2) Test</u>

An alternative approach to that described above is to work solely with hypothesis H_1 and to reject H_1 whenever the zero vector of parameters for C(z) falls outside a confidence region for θ_c under H_1.

Given an n component random variable θ distributed $N(\mathring{\theta}, \Sigma)$ we may define the random variable ξ by $\xi = (\theta-\mathring{\theta})^T \Sigma^{-1} (\theta-\mathring{\theta})$. ξ will then have the

χ^2 distribution with n degrees of freedom. Since $\hat{\theta}^N$ is asymptotically normally distributed $N(\overset{\circ}{\theta}, -L_{\theta\theta}^{-1}(\overset{\circ}{\theta}))$ we may formulate the following test:

Compute $\xi_c = \hat{\theta}_c^{NT} \Sigma_c^{-1} \hat{\theta}_c^N$ where $\hat{\theta}_c^N$ is the estimate of the components of Θ belonging to $C(z)$ and Σ_c is the submatrix of the estimate of $-L_{\theta\theta}^{-1}(\overset{\circ}{\theta})$ corresponding to θ_c. Then the decision procedure is:

Reject H_1 if $\xi_c < \gamma$

Accept H_1 if $\xi_c \geq \gamma$

where Prob $(\xi_c > \gamma | H_0) = \alpha$ and α is the prespecified risk level.

4.3 Examples.

To demonstrate the application of the ideas introduced in the previous sections we present two examples. The first one uses artificially generated data while the second involves recorded gross domestic product (GDP) and unemployment (UN) time series for the United Kingdom. In both examples the maximum likelihood estimates of the autoregressive moving-average models were obtained by minimizing (4.6) with respect to $\hat{\theta}$ over a specified compact set using Rosenbrock's hill climbing technique. The derivative of $V(\hat{\theta})$ was calculated by differentiating $V(\hat{\theta})$ analytically with respect to $\hat{\theta}$. The covariance $-L_{\theta\theta}^{-1}(\overset{\circ}{\theta})$ of the parameter estimate was computed using the well known identity

$$E[L_{\theta\theta}^N(\zeta^N,\theta)] = E[L_\theta^N(\zeta^N,\overset{\circ}{\theta}) L_\theta^{N^T}(\zeta^N,\overset{\circ}{\theta})] \tag{4.12}$$

between $L_{\theta\theta}^N(\zeta^N,\theta)$ and $L_\theta^N(\zeta^N,\overset{\circ}{\theta})$. As a result we use

$$\frac{1}{N} \sum_{i=1}^{N} L_\theta^i(\zeta^i,\hat{\theta}^N) \, L_\theta^{i^T}(\zeta^i,\hat{\theta}^N)$$

to estimate $L_{\theta\theta}^{N}(\zeta^{N},\overset{\circ}{\theta})$, where $L_{\theta}^{i}(\zeta^{i},\hat{\theta}^{N})$, $1 \leq i \leq N$ is evaluated using $V_{\theta}(\hat{\theta}^{N})$ computed for $1 \leq i \leq N$.

Example 1 200 pairs of univariate input and output observations were generated using the model

$$y = A\varepsilon_1 + Bu \tag{4.13}$$

$$u = \varepsilon_2$$

where

$$A(z) = \frac{1 + 0.4z^{-1}}{1 + 0.6z^{-1}} \quad , \quad B(z) = \frac{0.7z^{-1}}{1 + 0.9z^{-1}}$$

and ε_1, ε_2 were serially and mutually independent gaussian random variables generated by a standard computer subroutine and having distributions $N(0,1^2)$ and $N(0,0.5^2)$ respectively. For this sample the emperical distribution of these random variables was $N(0,1.047^2)$ and $N(0,0.496^2)$ respectively.

Since, in practice, the structure of $\Phi(z)$ is not known a-priori, the procedure of detecting feedback is performed using the following steps. First, find the most acceptable model using the likelihood ratio test (LRT) proposed in Section 4.2 for the case where feedback is believed to exist. In such a case, there will be no zero entry in $\Phi(z)$. Second, choose the most acceptable model using the LRT for the case where $\Phi(z)$ is upper block triangular. Third, use the LRT and χ^2-test to decide whether to accept or reject the alternative hypotheses.

In this example, the most acceptable identified model where feedback was assumed to exist is shown below with the estimated standard deviations of the estimated parameters shown in brackets.

$$\begin{bmatrix} y \\ u \end{bmatrix} = \begin{bmatrix} \dfrac{\underset{(\pm 0.157)}{1 + 0.414\ z^{-1}}}{\underset{(\pm 0.138)}{1 + 0.591\ z^{-1}}} & \dfrac{\underset{(\pm 0.232)}{0.723\ z^{-1}}}{\underset{(\pm 0.292)}{1 + 0.913\ z^{-1}}} \\ \underset{(\pm 0.032)}{-\ 0.014\ z^{-1}} & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix} \tag{4.14}$$

and

$$\text{Cov} \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix} = \begin{bmatrix} 1.047^2 & -0.015 \\ -0.015 & 0.496^2 \end{bmatrix} \tag{4.15}$$

Next we assumed $\Phi(z)$ was feedback free and the identified model was as follows:

$$\begin{bmatrix} y \\ u \end{bmatrix} = \begin{bmatrix} \dfrac{\underset{(\pm 0.309)}{1 + 0.429\ z^{-1}}}{\underset{(\pm 0.269)}{1 + 0.610\ z^{-1}}} & \dfrac{\underset{(\pm 0.101)}{0.722^{-1}}}{\underset{(\pm 0.018)}{1 + 0.913\ z^{-1}}} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix} \tag{4.16}$$

with

$$\text{Cov} \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix} = \begin{bmatrix} 1.044^2 & -0.015 \\ -0.015 & 0.496^2 \end{bmatrix} \tag{4.17}$$

It seems reasonable from inspection of (4.14) that the observed point process was feedback free. The hypothesis testing procedures applied to

1. $t = \frac{N-n_c}{n_c-n_0} \quad \frac{V(\theta_0) - V(\theta_c)}{V(\theta_c)}$

$$= \frac{200 - 8}{8 - 7} \times \frac{0.544 - 0.541}{0.541} = 1.065$$

Now the 5% risk level for $F(192,1)$ is 3.95 and so we accept the hypothesis H_0 at the 5% level.

2. In this example the inversion of the $(n_c-n_0) \times (n_c-n_0)$ submatrix Σ_c is simple since it is merely the scalar $(0.032)^2$. This gives

$$\xi = (0-\hat{\theta}_c)\ \hat{\Sigma}_c^{-1}\ (0-\hat{\theta}_c)$$

$$= (0.014)^2\ (0.032)^{-2} = 0.191$$

Since the 95% confidence region for θ_c under H_1 is given by $\xi = (\theta-\hat{\theta}_c)$ $\bar{\Sigma}^{-1}(\theta-\hat{\theta}_c) \leq 3.84$ we see that we reject H_1 at the 5% risk level by the χ^2 test also.

Example 2 65 values of the gross domestic product and unemployment time series for the United Kingdom from the first quarter of 1971 (1971 I) were used in this experiment. The G.D.P. data was at 1963 factor cost and seasonally adjusted; data for 1955-1967 is that quoted by Bray [35] from Treasury sources and that for 1968-1971 from Economic Trends, July 1971, Table 4 p. xii, col. 6. Unemployment denotes wholly unemployed, excluding school leavers and seasonally adjusted; 1955-1970 II is that quoted by Bray, while 1970 III-1971 I is from Trade and Industry, No. 28 October, p. 200.

Sixty-four pairs of normalized difference data were generated by computing $D_K = \frac{d_K - d_{K-1}}{d_K}$, k=1,...,65 for [GDP] = $\{GDP_K$, k=1,...,65$\}$ and

$[UN] = \{ UN_K, k=1,\ldots,65\}$. Further by subtracting form each series its average value over the sample we obtain two zero mean processes.

As before, we first assumed that $[GDP,UN]'$ is a process containing feedback. The most acceptable identified model was

$$\begin{bmatrix} GDP \\ UN \end{bmatrix} = \begin{bmatrix} \dfrac{1}{1 - \underset{(\pm 0.073)}{0.337}\, z^{-1}} & \dfrac{- \overset{(\pm 0.315)}{1.973}\, z^{-1} - \overset{(\pm 0.079)}{3.182}\, z^{-2}}{1 - \underset{(\pm 0.331)}{0.098}\, z^{-1} - \underset{(\pm 0.087)}{0.355}\, z^{-2} + \underset{(\pm 0.182)}{0.113}\, z^{-3}} \\ - \underset{(\pm 0.069)}{0.018}\, z^{-1} & \dfrac{1}{1 + \underset{(\pm 0.119)}{0.178}\, z^{-1}} \end{bmatrix} \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix} \tag{4.18}$$

$$\mathrm{Cov} \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix} = \begin{bmatrix} 5.848^2 & -2.901 \\ -2.901 & 1.054^2 \end{bmatrix} \tag{4.19}$$

Second we assumed the joint observed process is feedback free. This yielded the model:

$$\begin{bmatrix} GDP \\ UN \end{bmatrix} = \begin{bmatrix} \dfrac{1}{1 - \underset{(\pm 0.284)}{0.312}\, z^{-1}} & \dfrac{- \overset{(\pm 0.069)}{2.013}\, z^{-1} - \overset{(\pm 0.313)}{3.197}\, z^{-2}}{1 - \underset{(\pm 0.079)}{0.104}\, z^{-1} - \underset{(\pm 0.134)}{0.332}\, z^{-2} + \underset{(\pm 0.065)}{0.117}\, z^{-3}} \\ 0 & \dfrac{1}{1 + \underset{(\pm 0.094)}{0.116}\, z^{-1}} \end{bmatrix} \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix} \tag{4.20}$$

$$\mathrm{Cov} \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix} = \begin{bmatrix} 5.912^2 & -2.948 \\ -2.948 & 1.052^2 \end{bmatrix} \tag{4.21}$$

The hypothesis testing procedures yielded the following results

1. $$t = \frac{N-n_c}{n_c-n_0} \; \frac{V(\theta_0)-V(\theta_c)}{V(\theta_c)}$$

$$= 56 \times \frac{29.964 - 29.555}{29.555} = 0.775$$

At the 5% risk level H_0 is accepted for $t < F((64-8),(8-7)) = 4.02$. As a consequence the t test, or equivalently the likelihood ratio test, concludes that the process is feedback free at the 5% risk level.

2. For the χ^2 test we reject H_1 at the 5% risk level if $\theta_c = 0$ is in the 95% confidence region under H_1. This yields:

$$\text{Reject } H_1 \text{ if } \xi = (0-\hat{\theta}_i)\bar{\Sigma}^{-1}(0-\hat{\theta}_i) < 3.84$$

Substituting from (4.18) gives $\xi = (0.018)^2\,(0.069)^{-2} = 0.068$, and so H_1 is rejected at the 5% risk level.

In summary we conclude from both tests that the joint gross domestic product-unemployment series is feedback free.

In the light of Property 4 of Section 2 we may proceed to identify the relation between the input process [GDP] and the output process [UN]. From [36] this is known to be

$$[UN] = \frac{\underset{(\pm 0.5457)}{-2.3275} \quad \underset{(\pm 0.6124)}{+\,1.6396\, z^{-1}}}{1-\underset{(\pm 0.0773)}{1.7094\, z^{-1}} \quad \underset{(\pm 0.0658)}{+\,0.8074\, z^{-2}}}\,[GDP] + \frac{5.560}{1-\underset{(\pm 0.1268)}{0.224\, z^{-1}}}\,[E]$$

where [E] denotes an N(0,1) Gaussian noise process.*

*The standard deviation of the numerator of the noise transfer function is not recorded in [36].

Conclusion

In this paper we have proposed a formal definition for the idea of feedback between stationary stochastic processes. Our claim is that the special structures (2.5) for the canonical representation of the joint process and (2.7) for its spectrum correspond to the absence of feedback. We believe this claim is supported by the discussion in the first part of Section 2 and the consequences of the definitions given in the second part of Section 2. In Section 3 we showed that there are some intriguing consequences for filtering theory when an ordered pair of process is feedback free. From the point of view of applications Property 5 in Section 2 is the most important result of the definitions. It asserts that if (and only if) an ordered pair of observed processes is feedback free then there exists a unique representation of the first process in terms of the second process and a distrubance independent of the second process. We have illustrated these ideas by applications to simulated data and to econometric data from the United Kingdom. We believe there is great scope for the extension of these techniques in many practical applications.

Acknowledgements

The authors would like to thank K. D. Wall and R. L. Kashyap for their useful comments on an earlier draft of this paper.

References

[1] N. Wiener "The theory of prediction," Modern Mathematics for Engineers (Series 1) Chapter 8, ed. E. F. Beckenback, McGraw Hill, 1956.

[2] C. W. J. Granger, "Economic Processes Involving Feedback," Information and Control, Vol. 6, pp. 28-48, 1963.

[3] C. W. J. Granger and M. Hatanka, "Spectral Analysis of Economic Time Series," Princeton University Press, Princeton, New Jersey, 1964.

[4] W. Gersch, "Causality or Driving in Electrophysiological Signal Analysis," Math. Biosciences, Vol. 14, 177-196, 1972.

[5] W. Gersch, A. Midkiff, B. Tharp, "Alpha Rhythm Gemerators," Proc. Fifth International Conference in System Sciences, Computers in Biomedicine Supplement, Honolulu, Hawaii, January 1972.

[6] C-W Chan, "The Identification of Closed Loop Systems with Application to Econometric Problems," M.Sc. Dissertation, University of Manchester Institute of Science and Technology, 1972.

[7] G. A. Sims, "Money, Income and Causality," American Economic Review Vol. 62, 1972 pp. 540-552.

[8] D. Goodhart, D. Williams, and D. Gowland, "Money, Income and Causality: The U.K. Experience," submitted to American Economic Review.

[9] K. D. Wall, "An Application of Simultaneous Estimation to the Determination of Causality between Money and Income," unpublished report, Programme for Research into Econometric Methods, University of London, March 1974.

[10] H. Akaike, "On the Use of a Linear Model for the Identification of Feedback Systems," Ann. Inst. Statist. Math., Vol. 20, No. 3, 1968.

[11] T. Bohlin, "On the Problem of Ambiguities in Maximum Likelihood Identification," Automatica, Vol. 7, pp. 199-210, 1971.

[12] N. Wiener, P. Masani, "The Prediction Theory of Multivariate Stochastic Processes," I, Acta Mathematica 98, pp. 111-150, 1957.

[13] J. L. Doob, "Stochastic Processes," John Wiley, New York, 1953.

[14] D. C. Youla, "On the Factorization of Rational Matrices," IRE Trans. Information Theory, July 1961 IT-7(3), July 1961 pp. 172-189.

[15] W. B. Davenport and W. L. Root, "An Introduction to the Theory of Random Signals and Noise," McGraw-Hill, 1958.

[16] P. Whittle, "Prediction and Regulation," Van Nostrand, 1963.

[17] J. Rissanen, "An Algebraic Approach to the Problems of Linear Prediction and Identification," IBM Research Report (unpublished) October 1967.

[18] J. Rissanen and L. Barbosa, "Properties of Infinite Covariance Matrices and Stability of Optimum Predictors," Information Sciences Vol. 1 (1969), pp. 221-236.

[19] E. I. Jury, "Theory and Applications of the z-Transform Method," John Wiley, 1964.

[20] I. M. Gelfand and A. M. Yaglom, "Calculation of the Amount of Information about a Random Function Contained in Another Such Function," Uspekhi Mat. Nauk, Vol. 12, 1957, (Russian); English translation, Am. Math. Soc. Trans. (2), Vol. 12, pp. 199-246, 1959.

[21] K. J. Åström, T. Bohlin, S. Wensmark, "Automatic Construction of Linear Stochastic Dynamic Models for Stationary Industrial Processes with Random Distrubances Using Operating Records," Rep. TP 18.150, IBM Nordic Laboratories, Lindingö, Sweden, 1965.

[22] K. J. Aström and T. Bohlin, "Numerical Identification of Linear Dynamic Systems from Normal Operating Records," Proc. IFAC Symp. On the Theory of Self-Adaptive Control Systems 1965 Plenum Press, 1966.

[23] T. Bohlin, "On the Problem of Ambiguities in Maximum Likelihood Identification," Automatica, Vol. 7, pp. 199-210, 1971.

[24] L. Ljung, "On Consistency for Prediction Error Identification Methods," Division of Automatic Control, Lund Institute of Technology, Report 7605, March 1974.

[25] M. B. Priestley, "Estimation of Transfer Functions in Closed Loop Stochastic Systems," Automatica, Vol. 5, pp. 623-632, 1969.

[26] H. Akaike, "Some Problems in the Application of the Cross Spectral Method," Advanced Seminar on Spectral Analysis of Time Series (ed.) Harris, Wiley, New York, 1967.

[27] R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems," J. Basic Eng., ASME Vol. 82, pp. 35-45, 1960.

[28] A. Wald, "Note on the Consistency of the Maximum Likelihood Estimate," Ann. Math. Stat., Vol. 20, pp. 595-601, 1949.

[29] S. S. Wilks, "Mathematical Statistics," John Wiley, 1962.

[30] M. G. Kendall, A. Stuart, The Advanced Theory of Statistics, Vol. 2, Hafner Publishing Co., New York, 1967.

[31] P. E. Caines and J. Rissanen, "Maximum Likelihood Estimation of Parameters in Multivariate Gaussian Stochastic Processes," IEEE Trans. Information Theory, Vol. IT-20, No. 1, January 1974.

[32] J. Rissanen and P. E. Caines, "Estimation of Parameters in Stationary Gaussian Stochastic Processes," Submitted to Annals of Statistics.

[33] V. M. Popov, "Some Properties of the Control Systems with Irreducible Matrix-Transfer Functions," Lecture Notes in Math., No. 144, Springer-Verlag, Berlin, 1969.

[34] K. J. Åström, "On the Achievable Accuracy in Identification Problems," IFAC Symposium on Identification in Automatic Control Systems, Prague, Czechoslovakia, 1967.

[35] J. Bray, "Dynamic Equations for Econometric Forecasting with the G.D.P. - Unemployment Relation and the Growth of G.D.P. in the U.K. as an Example," Jnl. Royal Stat. Soc., Series A, Vol. 134, Part 2, 1971.

[36] K. D. Wall, A. J. Preston, J. W. Bray and M. H. Pestan, "Estimates for a Simple Control Model of the U.K. Economy," Proc. SSRC Conference, London, July 1972. In "Modelling the Economy," published by London Graduate School of Business, 1974.