SELECTION OF VARIABLES FOR IDENTIFICATION OF TRANSFER FUNCTION MODELS ON SAS SOFTWARE

James A. Deddens, University of Cincinnati
Wang Ping, Northern Jioatong University

INTRODUCTION. The purpose of this report is to discuss some of the complexities involved when modeling transfer function relationships between non-stationary time series using the SAS-ETS software. We will restrict our discussion to the case of only one input, and we will be particularly interested in the case where the input and output time series require different orders of differencing in order to achieve stationarity. Such relationships occur often with engineering control or economic data. The selection of the series of interest is somewhat arbitrary. For example, if one was interested in rice production, then one might select the monthly increase in production, the monthly production, or the cumulative production as the input or output series. PROC ARIMA provides great flexibility in modeling both the input and output variables. The input variable is called the Crosscorrelation variable (CC), and the output variable is called the Identification Variable (ID). It is important to correctly specify these two variables, since incorrect specification can give misleading information on the transfer relationship.

Recall that two time series X_t (the input) and Y_t (the output) satisfy a transfer function relationship (see [2, Chap. 1]) if

$$x_t = \phi(B)\, a_t$$

$$y_t = V(B)\, x_t + \psi(B)\, b_t$$

where a_t, b_t are independent white noise time series, and x_t, y_t represent the original series X_t, Y_t or some differenced form of X_t, Y_t. Often $s_t = V(B)\, x_t$ is called the signal, while $n_t = \psi(B)\, b_t$ is called the noise. The impulse response function can be written:

$$V(B) = \textstyle\sum_{k\geq 0} V_k B^k = B^b\, W(B)/S(B)$$

where b is called the lag, and W(B), S(B) are polynomials in the shift operator B. When both time series are stationary the usual "prewhitening" method of identification as recommended by Box-Jenkins consists of the following steps:

1) A univariate analysis of x_t, that is, identification and estimation of $\phi(B)$ using the autocorrelation function.

2) Identification and estimation of the transfer function V(B). This is accomplished by crosscorrelating the "prewhitened" output series $\beta_t = \phi B^{-1} y_t$ with the "prewhitened" input series $\alpha_t = \phi(B)^{-1} x_t$. The resulting crosscovariance function then equals $V_k \sigma_\alpha^2$. Hence the crosscorrelation function $\rho_{\alpha\beta}(k) = V_k \sigma_\alpha/\sigma_\beta$. Thus after prewhitening, the crosscorrelation function is proportional to the impulse response function (see section 11.2.1 in [2]). If the sample crosscorrelation function is not significantly different from 0 for negative k, then transfer function relationship exists and b, W(B), and S(B) can be identified and estimated using the sample crosscorrelation.

3) A univariate analysis of the residuals $y_t - V(B)\, x_t$ to determine the form of $\psi(B)$. (See Section 11.2.3 in [2]).

In case both time series are nonstationary but with the same order of differencing necessary to achieve stationarity, then one simply applies the above rules to the differenced series.

When the input and output series require different orders of differencing for achieving stationarity, then the situation is somewhat more complicated. In fact in this case, part of the problem seems to be to determine what is meant by the transfer function V(B). The question is whether or not to include some difference terms in V(B). The SAS/ETS software requires that the transfer function relationship be formulated in such a way so that the noise term is stationary in order to fully estimate the model. This will restrict the choice of the impulse response function.

The complexities will be illustrated by means of two simulated examples.

<u>EXAMPLE</u> 1. Consider the following example:

$$(1-B)X_t = (1-.5B)(1-.7B^{12})a_t$$

$$(1-B^{12})Y_t = 2B^3(1-.7B)(1-B^{12})X_t + b_t/(1+.8B)$$

or

$$Y_t = 2B3(1-.7B)\, X_t + b_t/(1+.8B)(1-B^{12})$$

where a_t, b_t are independent white noise time series. Notice that both X_t and Y_t are nonstationary. The time series X_t requires a single difference (1-B) in order to achieve stationarity, while the series Y_t requires a seasonal difference $(1-B^{12})$ in order to achieve stationarity. In this case, the transfer function would be

$$V(B) = 2B^3(1-.7B)(1-B^{12}),$$

or perhaps

$$V(B) = 2B^3(1-.7B)$$

if the difference term is not included. The former representation has a stationary noise term, whereas the latter has a nonstationary noise term.

In order to identify and estimate the input series X_t the usual Box-Jenkins method is used. This is accomplished with the statements:

```
IDENTIFY   VAR=X(1) ;
ESTIMATE Q=(1)(12)  NOINT ;
```

In order to identify the transfer function we need to specify the crosscorrelation variable (CC) and identification variables (ID). Given two time series X and Y with the above degrees of nonstationarity an initial choice might be:

```
IDENTIFY VAR=Y(12)  CROSSCOR = X(1) ;
```

To understand what SAS/ETS does with this specification, we need to quote the SAS/ETS manual page 80 in [1]:

> "The differences used in the prewhitening transformation, if any, are those differences specified in the model for the input variable and not specified by the CROSSCOR=".

This means that the theoretical prewhitening filter in this case is:

$$F = \text{Prewhitening Filter}$$
$$= 1/(1-.5B)(1-.7B^{12})$$

since a single difference is specified in both the model and CROSSCOR= statements for the input X. For illustrational purposes we will use the theoretical rather than the estimated prewhitening filter. This transforms the input variable into white noise:

$$\alpha_t = F\ X(1) = a_t$$

although the transformed output is:

$$\beta_t = F\ Y(12)$$
$$= 2B^3(1-.7B)(1-B^{12})\ F\ X + e_t$$
$$= 2B^3(1-.7B)(1-B^{12})a_{t/(1-B)}+e_t$$
$$= 2B^3(1-.7B)(1+B+B^2+\ldots+B^{11})a_t+e_t$$
$$= 2B^3(1+.3B+.3B^2+\ldots-.7B^{12})a_t+e_t$$

Here $e_t = F\ b_t/(1+.8B)$ is independent of a_t.

Thus the crosscovariance equals 2 if k=3, .6 if 3<k<15, -1.4 if k=15, and 0 otherwise. Hence the crosscorrelation function $\rho_{\alpha\beta}(k)$ between the prewhitened series α_t, β_t will have nonzero terms at t=3,4,5,6,7,B,9,10,11,12,13,14,15 and will not correspond to the transfer function V(B). Instead one obtains a "spread out" picture of the correct transfer function. Figure 1 shows such a crosscorrelation function for a simulation of 500 time points of the above model. The 500 time points were generated in a fashion similar to the example on page 140 of [1]. See Table 1. Notice that one can still conclude that a transfer function relationship exists (crosscorrelations for negative t values are not significant) and that the lag is 3 units.

Next consider the specification:

```
IDENTIFY  VAR=Y(12)  CROSSCOR=X ;
```

In this case, the theoretical prewhitening filter will be:

$$F = (1-B)/(1-.5B)(1-.7B^{12})$$

Note that the prewhitening filter contains a difference (1-B) since it was specified in the model statement but not in the Crosscorr= statement. This transforms the input variable into white noise:

$$\alpha_t = F\ X = a_t$$

and the output into:

$$\beta_t = F\ Y(12)$$
$$= 2B^3(1-.7B)(1-B^{12})\ F\ X + e_t$$
$$= 2B^3(1-.7B)(1-B^{12})\ a_t + e_t.$$

Here again $e_t = F\ b_t/(1+8B)$ is independent of a_t. Thus the crosscovariance equals 2 if k=3, -1.4 if k=4, -2 if k=15, 1.4 if k=16, and 0 otherwise. So the crosscorrelation function $\rho_{\alpha\beta}(k)$ between the prewhitened series will have nonzero terms at k=3,4,15,16 and will correspond to the transfer function:

$$V(B) = 2B^3(1-.7B)(1-B^{12})$$
$$= 2B^3-1.4B^4-2B^{15} + 1.4B^{16}$$

That is, $\rho_{\alpha\beta}(k) = V_k\ \sigma_\alpha/\sigma_\beta$. Figure 2 shows such a crosscorrelation function for a simulation of 500 time points. In order to estimate this model the statement would be:

```
ESTIMATE INPUT = (3$(1,12,13)X) NOINT ;
```

Table 2 contains the estimation results for the simulation of 500 time points. One can see that the polynomial contains a factor of the form $(1-B^{12})$. This can also be seen from the crosscorrelation function, since the spikes at 15, 16 are roughly equal in magnitude with opposite sign than the spikes at 3, 4. Thus one could use the statement

```
ESTIMATE INPUT = (3$(1)(12)X) P=1 NOINT ;
```

Recognizing that there is a factor of the form $(1-B^{12})$ in the transfer function one can incorporate this information by using the statements:

```
IDENTIFY  VAR=Y(12)  CROSSCOR=X(12) ;
```

The difficulty with this specification is that the crosscorrelation of the prewhitened series is related to the transfer function in a complicated manner. In order to see this, note that the theoretical prewhitening filter in this case is:

$$F = (1-B)/(1-.5B)(1-.7B^{12})$$

But this filter does not transform the input series into white noise. In fact, the "prewhitened" input is:

$$\begin{aligned} \alpha_t &= F\ X(12) \\ &= (1-B^{12})\ F\ x_t \\ &= (1-B^{12})\ a_t \\ &= a_t - a_{t-12} \end{aligned}$$

while as before the prewhitened output becomes:

$$\begin{aligned} \beta_t &= F\ Y(12) \\ &= 2B^3(1-.7B)(1-B^{12})a_t + e_t \end{aligned}$$

Thus we have overdifferenced the input, and it is no longer white noise. Theoretically the crosscorrelation function between the "prewhitened" pair α_t and β_t will be

$$\begin{aligned} \rho_{\alpha\beta}(k) &= \mathrm{cov}(\alpha_{t-k}, \beta_t) / \sigma_\alpha \sigma_\beta \\ &= \frac{(\mathrm{cov}(a_{t-k}, \beta_t) - \mathrm{cov}(a_{t-12-k}, \beta_t))}{\sqrt{2}\ \sigma_a\ \sigma_\beta} \end{aligned}$$

Since $\mathrm{cov}(a_{t-k}, \beta_t)$ is nonzero when $k=3,4,15,16$, the correlation $\rho_{\alpha\beta}(k)$ will have nonzero terms at $k=-9,-8,3,4,15,16$. Figure 3 shows such a crosscorrelation function for 500 time points. Notice that this could mislead one into thinking that a feedback relationship exists, since there are significant crosscorrelations at $k=-9,-8$. In general, the addition of an extraneous factor $(1-B^s)$ into the CROSSCOR= statement will introduce extra terms into the crosscorrelation function at k-s. The proper estimation statement now becomes:

```
IDENTIFY  VAR=Y(12)  CROSSCOR=(X(12)) ;
ESTIMATE  INPUT=(3$(1) X) P=1 NOINT ;
```

Table 3 gives the estimation results f r the simulation of 500 time points.

Thus we see that for this example one actually needs to do two identification procedures. First

```
IDENTIFY  VAR=Y(12)  CROSSCOR=X
```

to identify the transfer function model, and then

```
IDENTIFY  VAR=Y(12)  CROSSCOR=X(12)
```

to estimate the model parameters. However, the crosscorrelation function for the second identification step is ignored. One only does the identification because it is necessary in order to perform the desired estimation.

In this example, no specification in SAS/ETS will yield a crosscorrelation function proportional to the impulse response function:

$$2B^3(1-.7B)$$

One might suspect the statements

```
IDENTIFY  VAR = Y  CROSSCOR = (X)  ;
```

will yield the desired function. However, the extremely large variance of the nonstationary prewhitened output β_t causes the sample crosscorrelation function to have no significantly nonzero terms.

EXAMPLE 2. In order to illustrate another type of complexity, consider the following example:

$$(1-B)X_t=(1-.5B)(1-.7B^{12})a_t$$

$$(1-B)Y_t=2B^3(1+.7B)X_t + b_t/(1+.8B)$$

Here again a_t, b_t are independent white noise time series. X_t requires a first difference $(1-B)$, while Y_t requires a second difference $(1-B)(1-B)$ in order to achieve stationarity. In this case the inpulse response function is

$$V(B) = 2B3(1+.7B).$$

As in Example 1, the statements for the input are:

```
IDENTIFY VAR=X(1) ;
ESTIMATE Q=(1)(12) NOINT ;
```

Given the order of nonstationarity of X and Y, identification of the transfer function begins with the statements

Given the order of nonstationarity of X and Y, identification of the transfer function begins with the statements:

```
IDENTIFY VAR=Y(1,1)  CROSSCOR=(X(1)) ;
```

resulting in the theoretical prewhitening filter:

$$F = 1/(1-.5B)(1-.7B^{12})$$

so that the prewhitened input

$$\alpha_t = F\ X(1) = a_t$$

is white noise, while the prewhitened output is

$$\beta_t = F\ Y(1,1)$$
$$= 2B^3(1+.7B)\ F\ X(1) + e_t$$
$$= 2B^3(1+.7B)\ a_t + e_t$$

Here $e_t = F(1-B)b_t/(1+.8B)$ is independent of a_t. Hence the crosscovariance of the prewhitened series is equal to 2 if k=3, -1.4 if k=4 and 0 otherwise. Thus the crosscorrelation is proportional to the impulse response function V(B). Figure 4 shows such a crosscorrelation function for a simulation of 500 time points. However, upon specifying

```
ESTIMATE INPUT=(3$(1)X) ;
```

one obtains that the noise term is noninvertible. In this case one needs to "cancel" a difference factor throughout. We are thus led to the statement:

```
IDENTIFY VAR=Y(1)  CROSSCOR=(X) ;
```

which results in the prewhitening filter:

$$F = (1-B)/(1-.5B)(1-.7B^{12})$$

and hence exactly the same prewhitened input and output as above, and hence the same crosscorrelation function. Finally the statement:

```
ESTIMATE INPUT=(3$(1)X) P=1 NOINT ;
```

allows one to correctly estimate the model with a stationary noise term. Table 4 contains the estimation results for the simulation of 500 time points.

<u>CONCLUSION</u>. We conclude by giving some guidelines that we have found helpful for the identification and estimation of transfer function relationships using SAS/ETS software.

1) Identify and estimate the input series X_t.

2) Select the ID as a stationary time series.

3) For identification purposes select the CC as either X_t or X_t differenced by some (but perhaps not all) of the difference terms needed to make X_t stationary.

Generally there will be many choices for the CC. SAS/ETS will always provide the correct prewhitening filter to make the prewhitened input series white noise. Therefore the crosscorrelation function between the prewhitened series will represent a correct form of the transfer function. Although different pairs of ID and CC will give different crosscorrelations, indicating different functions, they will be consistent, especially regarding the existence of a transfer function relationship and the corresponding lag. Some pairs will give a clearer picture of the relationship than others. For example, Figure 2 over Figure 1 in Example 1. Additionally there may be indications that certain factors correspond to difference factors. Finally since estimation within the SAS/ETS software requires that the noise term be stationary, occasionally redundant difference terms need to be cancelled from both the ID and CC terms, as in Example 2.

4) For estimation purposes it may be necessary to add difference factors to the CC term, or to delete some difference factors from the ID and CC terms, depending on the results of step 3).

In case difference factors are added to the CC variable in step 4), the crosscorrelation between the prewhitened series needs to be ignored since the prewhitened input may not be white noise.

We hope that by presenting some examples we have been able to illustrate the complexities of deciding how to choose the ID and CC variables when modeling transfer function models. Of course, when modeling real data, one does not know the model. However, the above guidelines should help one proceed in case one has no apriori model in mind.

The second author would like to express her thanks to the Department of Mathematical Sciences of the University of Cincinnati for their hospitality during her visit.

REFERENCES

1) SAS/ETS Users Guide, 1982 Edition, SAS Institute.

2) G. Box and G. Jenkins, Time Series Analysis: Forecasting and Control, Holden Day, 1976.

TABLE 1
SIMULATION PROGRAM
EXAMPLE 1

```
DATA SIMULATE;
 X=0; X1=0;....X16=0;
 Y=0; Y1=0;....Y12=0
 A1=0; A2=0;...A14=0;
 B1=0; M1=0; M2=0;
DO T=-100 TO 500 BY 1;
 A1=NORMAL(1234567);
 B1=NORMAL(2341235);
X=X1+A1-.5*A2-.7*A13+.35*A14;
Y=Y12+2*(X3-.7*X4-X15+.7*X16)+M1;
M1=-.8*M2+B1;
 Y12=Y11;...Y2=Y1;Y1=Y;
 X16=X15;...X2=X1;X1=X;
 A14=A13;...A2=A1;
 M2=M1;
 IF T>0 THEN OUTPUT; END;
```

FIGURE 1
CROSSCORRELATION--36 LAGS
ID=Y(12)----CC=X(1)

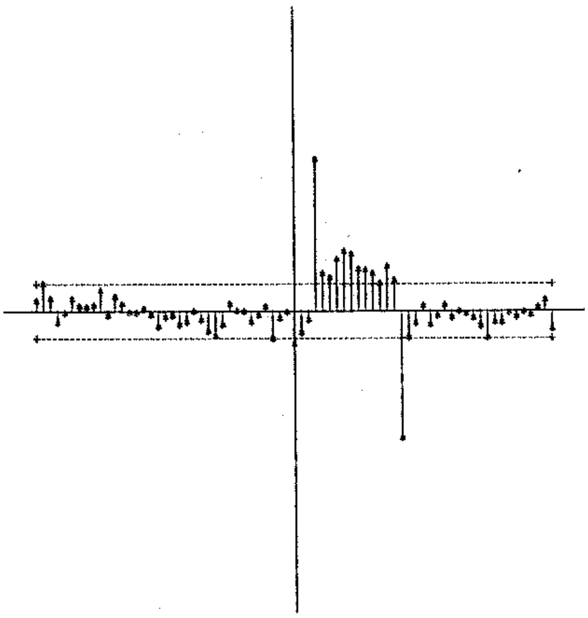

FIGURE 2
CROSSCORRELATION--36 LAGS
ID=Y(12)----CC=X

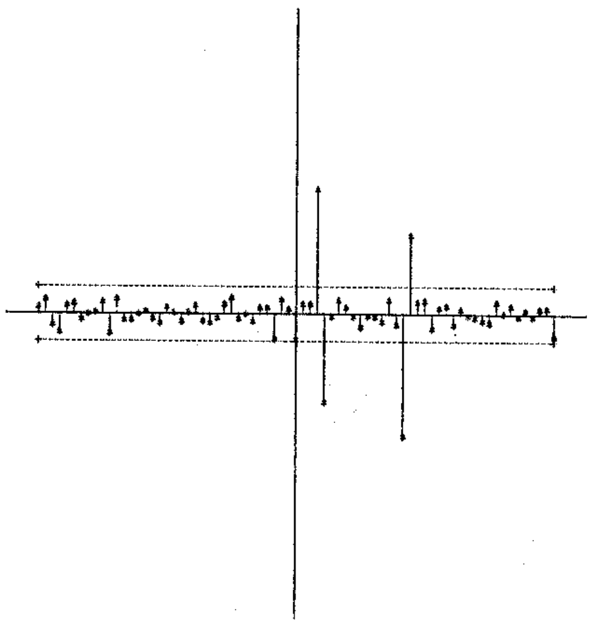

TABLE 2
EXAMPLE 1
ESTIMATION
VAR=Y(12)--CC=X

Parm	Est	Std E	Lag	Var
AR1,1	- .8144	.0266	1	Y
NUM1	1.9822	.0339	0	X
NUM1,1	1.3877	.0347	1	X
NUM1,2	2.0671	.0346	12	X
NUM1,3	-1.4920	.0338	13	X

Variance Est = .9735

FIGURE 3
CROSSCORRELATION--36 LAGS
ID=Y(12)-----CC=X(12)

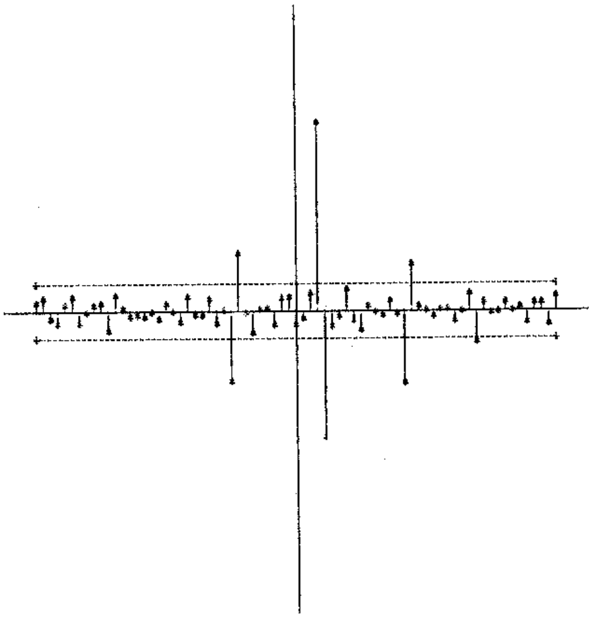

FIGURE 4--EXAMPLE 2
CROSSCORRELATION--36 LAGS
ID=Y(1,1)-----CC=X(1)

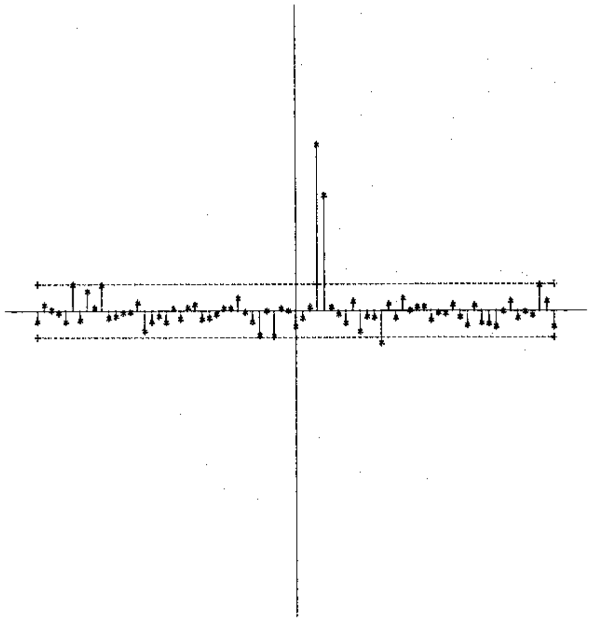

TABLE 3
EXAMPLE 1
ESTIMATION
VAR=Y(12)--CC=X(12)

Parm	Est	Std E	Lag	Var
AR1,1	-0.8117	0.0267	1	Y
NUM1	2.0227	0.0191	0	X
NUM1,1	1.4398	0.0191	1	X

Variance Est = .9539

TABLE 4
EXAMPLE 2
ESTIMATION
VAR=Y(1)---CC=X

Parm	Est	Std E	Lag	Var
AR1,1	-0.8045	0.0268	1	Y
NUM1	2.0172	0.0315	0	X
NUM1,1	1.4122	0.0315	1	X

Variance Est = .9820