Introduction to Transfer Functions

Consider explaining a response series $\{Y_t\}$ in terms of a single input series $\{X_t\}$.

Regression and Regression with ARMA Errors

The simplest possible model is a simple regression model:

$$Y_t = C + v_0 X_t + \varepsilon_t \, .$$

When fitting such a model, the estimated errors $\hat{\varepsilon}_t$ are often found to be serially correlated. In this case, we might model the errors ε_t as an ARMA(p, q) process (or perhaps an ARIMA process):

$$Y_t = C + v_0 X_t + N_t$$

where N_t denotes an ARIMA process.

Adding Lagged Terms in the Input Variable

Models may often be improved by adding lagged terms.

A transfer function model has the form

$$Y_t = C + v_0 X_t + v_1 X_{t-1} + v_2 X_{t-2} + \cdots + v_h X_{t-h} + N_t$$

where $\{N_t\}$ is an ARIMA process, and the series $\{X_t\}$ and $\{N_t\}$ are independent.

If we define $v(B) = v_0 + v_1 B + v_2 B^2 + \cdots + v_h B^h$, then

$$v(B)X_t = v_0 X_t + v_1 X_{t-1} + v_2 X_{t-2} + \cdots + v_h X_{t-h}$$

and the transfer function model can be written as

$$Y_t = C + v(B)X_t + N_t\,.$$

Remarks:

The coefficients $v_0, v_1, v_2, \ldots$ are often called the v-weights.

$v(B)$ is called the "transfer function" or "impulse response function" because ...

$v(B)$ "transfers" the changes in $\{X_t\}$ to $\{Y_t\}$.

In the simplest case where $Y_t = v(B)X_t$, if the input $\{X_t\}$ consists of a unit pulse at time zero (i.e. $X_0 = 1$ and $X_t = 0$ for $t \neq 0$), then the output is $Y_t = v_t$ for $t = 0, 1, 2, \ldots, h$ (and $Y_t = 0$ otherwise). In other words, the v-weights give the response (the output) when the input is an "impulse".

If we make no assumptions about the pattern of the v-weights (i.e. they are completely arbitrary), then the model

$$Y_t = C + v(B)X_t + N_t \, .$$

is called a **linear transfer function** because it has the form of a linear regression. It is also called a **free-form distributed lag model**.

(We soon discuss transfer function models which assume the v-weights follow certain patterns.)

N_t is often called the "noise process".

If N_t is an ARMA process, then we can write

$$N_t = \frac{\theta(B)}{\phi(B)} a_t$$

so that the transfer function model becomes

$$Y_t = C + v(B)X_t + \frac{\theta(B)}{\phi(B)} a_t \,.$$

Identifying Linear Transfer Function Models: A Simple Approach

Choose an initial h which is relatively large, and fit a multiple regression of Y_t on X_t, X_{t-1}, $\ldots$, X_{t-h}. Study the ACF/PACF of the residuals from this multiple regression to identify a plausible ARMA model for the noise process. Then re-fit the multiple regression, but now assuming ARMA errors. Use the p-values for the estimates $\hat{v}_i$ to decide which values of v_i are likely to be non-zero, and make a final choice of h.

This approach is easy to implement in SAS using PROC ARIMA.

A variation: In the first stage, instead of using a plain multiple regression, you can use regression with ARMA errors in which a simple ARMA model (called a proxy model) such as an AR(2) is temporarily assumed. This is used to obtain estimates $\hat{N}_t$ of the errors, which are then used to identify the final ARMA model for N_t. (This method is more work to implement in SAS.)

Models with a Non-Stationary Noise Process

Suppose the residuals $\hat{N}_t$ from the multiple regression model appear non-stationary, and are reasonably modeled by an ARIMA(p, d, q) model with $d = 1$ and no constant.

Then we assume $N_t \sim$ ARIMA$(p, 1, q)$, and fit a linear transfer function model with ARIMA$(p, 1, q)$ noise.

How do we do this?

$$\phi(B)(1-B)N_t = \theta(B)a_t \quad \Rightarrow \quad N_t = \frac{\theta(B)}{\phi(B)(1-B)}a_t$$

so that the transfer function model becomes

$$Y_t = C + v(B)X_t + \frac{\theta(B)}{\phi(B)(1-B)}a_t$$

or equivalently

$$(1-B)Y_t = v(B)(1-B)X_t + \frac{\theta(B)}{\phi(B)}a_t$$

upon multiplying both sides by $1-B$ and using $(1-B)C=0$.

This means we can fit the original linear transfer function model with ARIMA(p,1,q) errors by differencing both Y_t and X_t and fitting a model with ARMA(p,q) errors to the differenced data.

PROC ARIMA requires us to use this approach.

Rational Distributed Lag Models

A fairly general form for the transfer function is the rational polynomial:

$$v(B) = \frac{B^b \omega(B)}{\delta(B)}$$

$$\text{where } \omega(B) = \omega_0 - \omega_1 B - \omega_2 B^2 - \cdots - \omega_h B^h$$

$$\text{and } \delta(B) = 1 - \delta_1 B - \delta_2 B^2 - \cdots - \delta_r B^r \, ,$$

and $b \geq 0$ is called the delay, dead time, shift, or lead time. If $b = 0$, a change in X has an immediate effect on Y. But if $b > 0$, the effect is delayed by b time units.

If we expand $v(B)$ as a series:

$$v(B) = \sum_{i=0}^{\infty} v_i B^i$$

we get $v_i = 0$ for $i < b$. This means

$$v(B)X_t = \sum_{i=0}^{\infty} v_i X_{t-i} = v_b X_{t-b} + v_{b+1} X_{t-b-1} + \cdots$$

The Noise Process

If the noise process N_t is a mean zero ARMA(p,q) process, we may write

$$N_t = \frac{\theta(B)}{\phi(B)} a_t$$

$$\text{where } \phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$$
$$\text{and } \theta(B) = 1 - \theta_1 B - \cdots - \theta_q B^q$$

and a_t is a random shock sequence (white noise).

To estimate the transfer function model

$$Y_t = C + \frac{B^b \omega(B)}{\delta(B)} X_t + \frac{\theta(B)}{\phi(B)} a_t$$

in PROC ARIMA, we use

```
ESTIMATE INPUT=(b$(1,2,...,h)/(1,2,...,r)X) P=p Q=q ;
```

where b, r, h, p, q are integers denoting the delay, denominator order, numerator order, AR-order, and MA-order. If $b = 0$, omit `b$`. If $r = 0$ (i.e., $\delta(B) = 1$), omit `/(1,2,...,r)`. If $h = 0$, (i.e., $\omega(B) = 1$), omit `(1,2,...,h)`. The commas in `INPUT=...` are optional.

Actually, SAS allows $\omega(B)$, $\delta(B)$, $\theta(B)$, $\phi(B)$ to have arbitrary multiplicative forms, and you can specify exactly what lag terms you want in each factor. For example, to estimate the model:

$$Y_t = C + \frac{B^2(\omega_0 - \omega_{1,1}B^3)(1 - \omega_{2,1}B^{12})}{1 - \delta_{1,1}B^2 - \delta_{1,2}B^6} X_t + \frac{1 - \theta_{1,1}B^3 - \theta_{1,2}B^5 - \theta_{1,3}B^8}{(1 - \phi_{1,1}B - \phi_{1,2}B^2)(1 - \phi_{2,1}B^{12})} a_t$$

we use

```
ESTIMATE INPUT=(2$(3)(12)/(2,6)X) P=(1,2)(12) Q=(3,5,8) ;
```

The ALTPARM option

If the ALTPARM option (alternative parameterization) is used in the ESTIMATE statement, then the first numerator factor is taken to be

$$\omega_0(1 - \omega_1 B^1 - \cdots - \omega_h B^h) \quad \text{instead of} \quad \omega_0 - \omega_1 B^1 - \cdots - \omega_h B^h$$

in our first example, and

$$\omega_0(1 - \omega_{1,1} B^3) \quad \text{instead of} \quad \omega_0 - \omega_{1,1} B^3$$

in our second.

Identifying Transfer Functions

How do we identify transfer function models? That is, how do we choose the delay b and the orders of the numerator and denominator polynomials $\omega(B)$ and $\delta(B)$ in the transfer function $v(B) = B^b\omega(B)/\delta(B)$?

Suppose $\{X_t\}$ is a random input series.

Definition: jointly stationary processes

If $\{Y_t\}$ and $\{X_t\}$ are both stationary processes and the correlation $\text{Corr}(Y_s, X_t)$ depends on the time difference $s - t$ only, we say that $\{Y_t\}$ and $\{X_t\}$ are **jointly** stationary.

If $\{X_t\}$ and $\{Y_t\}$ are jointly stationary, then we can define the cross-correlation function:

$$\rho_{xy}(s) = \text{Corr}(x_t, y_{t+s}) = \text{Corr}(y_t, x_{t-s})$$

Identifying Transfer Functions for Jointly Stationary Series

Suppose $\{X_t\}$ and $\{Y_t\}$ are jointly stationary, and

$$Y_t = C + v(B)X_t + N_t$$

where the series $\{X_t\}$ and $\{N_t\}$ are independent.

We describe how to identify transfer functions in this special case.

We begin by discussing the even more special case where X_t is white noise so that $\text{Corr}(X_s, X_t) = 0$ when $s \neq t$.

Fact: If X_t is white noise, then

$$v_k = \frac{\sigma_y}{\sigma_x}\,\rho_{xy}(k)\,, \quad k = 0, 1, 2, 3, \ldots$$

so that the v-weights are proportional to the theoretical CCF (and will be approximately proportional to the sample CCF if the series length n is large enough).

Proof: Mean center each series: $\tilde{X}_t = X_t - \mu_x$ and $\tilde{Y}_t = Y_t - \mu_y$. Then the constant C disappears and our model becomes

$$\tilde{Y}_t = v(B)\tilde{X}_t + N_t\,.$$

Thus

$$\begin{aligned}
\rho_{xy}(k) &= \mathrm{Corr}(Y_t, X_{t-k}) = \frac{\mathrm{Cov}(Y_t, X_{t-k})}{\sigma_y \sigma_x} \\
&= \frac{1}{\sigma_y \sigma_x} E(\tilde{Y}_t \tilde{X}_{t-k}) \\
&= \frac{1}{\sigma_y \sigma_x} E\left[\tilde{X}_{t-k}(N_t + v_0 \tilde{X}_t + v_1 \tilde{X}_{t-1} + v_2 \tilde{X}_{t-2} + \cdots)\right] \qquad (\dagger) \\
&= \frac{1}{\sigma_y \sigma_x} (v_k \sigma_x^2) = \frac{\sigma_x}{\sigma_y} v_k \,.
\end{aligned}$$

In going from ($\dagger$) to the next line we used

$$E\tilde{X}_{t-k} N_t = E\tilde{X}_{t-k}\, EN_t = 0 \cdot 0 = 0$$

by independence of X_s and N_t for all s and t. We also used

$$E\tilde{X}_{t-k}^2 = \sigma_x^2 \quad \text{and} \quad E\tilde{X}_{t-k}\tilde{X}_{t-j} = 0 \quad \text{for } j \neq k.$$

Essentially same argument shows

Another Fact:

$$\rho_{xy}(k) = \text{Corr}(x_t, y_{t+k}) = 0 \quad \text{for } k = -1, -2, -3, \ldots$$

This says that, if $\{X_t\}$ and $\{Y_t\}$ satisfy a transfer function model and $\{X_t\}$ is white noise, the current and past values of Y cannot influence future values of X; there is no feedback.

What good is the special case we just proved?

Any jointly stationary series $\{Y_t\}$ and $\{X_t\}$ can be reduced to this case by **pre-whitening**:

- Find an ARMA model for X_t:

  $$\phi_x(B)X_t = \theta_x(B)b_t \, .$$

  Here b_t denotes a white noise series. This model gives us a filter $f(B)$ which converts X_t into white noise:

  $$b_t = \frac{\phi_x(B)}{\theta_x(B)}X_t = f(B)X_t \quad \text{where} \quad f(B) = \frac{\phi_x(B)}{\theta_x(B)} \, .$$

  Thus $f(B)X_t$ is white noise.
- Apply the filter $f(B)$ to both $\{X_t\}$ and $\{Y_t\}$. Let $X_t' = f(B)X_t$ and $Y_t' = f(B)Y_t$. The series X_t' (formerly called b_t) is white noise. This process is called pre-whitening.

- The series X_t' and Y_t' follow a transfer function model with the same transfer function $v(B)$ as the original model:

$$Y_t' = v(B)X_t' + N_t' \qquad (\ddagger)$$

since multiplying both sides of $Y_t = v(B)X_t + N_t$ by $f(B)$ gives

$$f(B)Y_t = v(B)f(B)X_t + f(B)N_t$$

which has the form in ($\ddagger$).
- The CCF between X_t' and Y_t' will be proportional to the v-weights. (This is exactly true for the theoretical CCF if we use the true model for X_t to do the pre-whitening, and approximately true for the sample CCF if the series are long enough and we find a reasonable model for X_t.)
- Use the sample CCF between X_t' and Y_t' to determine the form of $v(B)$.

Comment on Feedback: As shown earlier, if the pre-whitened series X_t' and Y_t' follow a transfer function model, then $\rho_{x'y'}(k) = 0$ for $k < 0$. If the sample CCF $\hat{\rho}_{x'y'}(k)$ is "large" at one or more negative lags, this is evidence of feedback and the model is suspect.

After identifying a tentative form for $v(B)$, use PROC ARIMA to fit this model (not trying to model the noise) and use the residuals to identify an ARMA process for the noise.

Fit and refine the resulting model.

In your final model, the residuals should resemble white noise and also be uncorrelated with the input series. Examine the usual residual diagnostics and also the cross-correlations between the residuals and the input series.

What if the series $\{X_t\}$ and $\{Y_t\}$ are not jointly stationary?

Try differencing.

If $\{X_t\}$ and $\{Y_t\}$ are not stationary, try to find orders of differencing (non-seasonal and seasonal) such that

$$X_t^* = (1-B)^{d_1}(1-B^s)^{D_1} X_t \quad \text{and} \quad Y_t^* = (1-B)^{d_2}(1-B^s)^{D_2} Y_t$$

are stationary.

If $\{X_t^*\}$ and $\{Y_t^*\}$ appear to be jointly stationary, then apply the pre-whitening procedure to identify a transfer function model and then identify an ARMA noise model (possibly seasonal, say, an ARIMA$(p,0,q)(P,0,Q)$) for these series. Estimating this model leads to:

$$Y_t^* = \frac{B^b \omega(B)}{\delta(B)} X_t^* + \frac{\theta(B)}{\phi(B)} a_t$$

If the orders of differencing are the same for the two series:

$$d_1 = d_2 \equiv d \quad \text{and} \quad D_1 = D_2 \equiv D$$

this model has a nice interpretation in terms of the original series:

$$(1-B)^d(1-B^s)^D Y_t = \frac{B^b\omega(B)}{\delta(B)}(1-B)^d(1-B^s)^D X_t + \frac{\theta(B)}{\phi(B)}a_t$$

becomes (upon dividing through by $(1-B)^d(1-B^s)^D$)

$$Y_t = \frac{B^b\omega(B)}{\delta(B)}X_t + \frac{\theta(B)}{(1-B)^d(1-B^s)^D\phi(B)}a_t\,.$$

We recognize this as a transfer function model with an ARIMA noise process:

$$N_t = \frac{\theta(B)}{(1-B)^d(1-B^s)^D\phi(B)}a_t$$

is an ARIMA$(p, d, q)(P, D, Q)$ process.