### Estimation

After identifying plausible ARMA models, estimate the parameters in a model using the ESTIMATE statement in PROC ARIMA:

- ESTIMATE P=3; fits an AR(3) model. (Similarly for any AR order.)
- ESTIMATE Q=2; fits an MA(2) model. (Similarly for any MA order.)
- ESTIMATE P=2 Q=1; fits an ARMA(2,1) model. (Similarly for any orders.)

- ESTIMATE P=(1,3,5,7); fits an AR model including terms only at lags 1, 3, 5, 7.
- ESTIMATE Q=(2,4); fits an MA model including terms only a lags 2, 4.
- ESTIMATE P=(1,3) Q=(2,4); fits a model with AR and MA terms only at the specified lags.

There are three methods of estimation available:

- CLS (conditional least squares, the default)
- ULS (unconditional least squares)
- ML (maximum likelihood)

The option METHOD=ML or METHOD=ULS in the ESTIMATE statement specifies use of these methods.

ML is the preferred method when the shocks  $a_t$  are independent and (approximately) normally distributed with constant variance.

ML estimates the parameters by those values which maximize the "likelihood" of the observed data  $z_1, z_2, \ldots, z_n$ , assuming the shocks  $a_t$  are independent  $N(0, \sigma_a^2)$ .

CLS estimates parameters by those values which minimize a certain "conditional" sum of squared errors. For an AR(p) model, CLS minimizes

$$\sum_{t=1}^{n} (\tilde{z}_t - \phi_1 \tilde{z}_{t-1} - \dots - \phi_p \tilde{z}_{t-p})^2$$

where  $\tilde{z}_t = z_t - \mu_z$  for  $t \ge 1$  and  $\tilde{z}_t$  is set equal to zero for t < 1.

#### Model Diagnostics

For each parameter, SAS reports an estimate, standard error, t-value, and p-value, with interpretations similar to the same quantities in regression.

$$t\text{-value} = \frac{\text{estimate}}{\text{standard error}} = \frac{\hat{\theta}}{\text{SE}(\hat{\theta})}$$

Here (and later) we use  $\theta$  to denote a "generic" parameter.

The p-value for a parameter reports the approximate probability of getting a t-value by chance whose magnitude is as large as that observed when the true value of the parameter is zero.

If the *p*-value for a parameter  $\theta$  is small (say, less than .05), we reject  $H_0: \theta = 0$ , and conclude that  $\theta$  is (probably) nonzero. The corresponding term (usually) improves the performance of the model and is retained in the model.

If the *p*-value is **not** small, we usually drop that term from the model. This yields a more "parsimonious" model (one with fewer parameters) which does about as well the original model.

The output also gives estimates of:

| Quantity     | Name in Output     |
|--------------|--------------------|
| $\mu_z$      | MU                 |
| C            | Constant Estimate  |
| $\sigma_a^2$ | Variance Estimate  |
| $\sigma_a$   | Std Error Estimate |

#### Residuals

SAS computes residuals  $\hat{a}_t$  which are estimates of the random shocks  $a_t$ .

The residuals should behave like random shocks.

• The sample ACF of the residuals should resemble that of random shocks. The "Autocorrelation

Check for Residuals" (the Ljung-Box test) should not be significant.

- The residuals should have (approximately) mean zero and constant variance. Look for these properties in the plot of residuals versus time order, and in the plot of residuals versus predicted values. In these plots, it is useful to add a smooth estimate of the mean using the LOESS smoother.
- The residuals should be approximately normally distributed, at least when you are using ML estimation. Judge this in the Q-Q Plot (residuals versus normal quantiles) which should be roughly a straight line, and the histogram/density plot.

The "predicted values" used in the plot of residuals versus predicted values are the one-step-ahead predictions described later. They are similar to the fitted values (also called predicted values) in regression. For a good model, the residuals should be independent of the predicted values; there should be no relationship between the residuals and predicted values.

## **Comparing Different Models**

If more than one model has reasonable residual diagnostics (or if all the models have problems), then one strategy is to choose the model with the smallest AIC or SBC value. • AIC is Akaike's Information Criterion:

$$AIC = -2\ln(L) + 2k$$

where L is the Likelihood value and k is the number of estimated parameters.

• SBC is *Schwarz's Bayesian Criterion*, also known as BIC for *Bayesian Information Criterion*:

$$SBC = -2\ln(L) + k\ln(n)$$

where n is the number of residuals.

# **MINIC Method**

The minimum information criterion (MINIC) method can tentatively identify the order of a stationary and invertible ARMA process.

This is available as an option in the IDENTIFY statement of PROC ARIMA.

For a range of values for p and q, a table supplies an estimate of the BIC (=SBC) divided by n (the series length) that would be obtained if an ARMA(p, q) model were fit to the data.

The model with the smallest estimated BIC could be used as a tentative model choice.

The models with the smallest estimated BIC's are worth considering.