Chapter 8: Model Diagnostics

- Model diagnostics involve checking how well the model fits.
- If the model fits poorly, we consider changing the specification of the model.
- A major tool of model diagnostics is *residual analysis*.
- We will also check *overparameterized models*.
- That is, we fit a slightly more general model than the one we originally specified.
- We can check to see whether we need the more general model, or whether the originally specified model is sufficient.

Residual Analysis

- We have seen residual analysis in Chapter 3 when we examined the residuals after fitting a trend model.
- For AR models, the definition of the residuals is straightforward.
- For example, with the $AR(1)$ model containing a constant term, $Y_t = \theta_0 + \phi Y_{t-1} + e_t$ the residuals are:

$$\hat{e}_t = Y_t - \hat{\theta}_0 - \hat{\phi} Y_{t-1}.$$

Residual Analysis in General ARMA Models

- For ARMA models, we use the infinite autoregressive representation of the model, whose estimated coefficients are functions of the estimated ϕ's and θ's.
- The residuals are calculated as $Y_t - \hat{Y}_t$, where $\hat{Y}_t$ is the best forecast of Y_t based on $Y_{t-1}, Y_{t-2}, \ldots$ (we will discuss this forecasting concept more in Chapter 9).
- In any case, if the model is correctly specified, the residuals should have the properties of white noise (independent normal r.v.'s with zero mean and common variances).
- If the residuals deviate from this white-noise behavior in some way, we may want to change our model to something more appropriate.

Residual Plots

- The most basic residual plot is the plot of standardized residuals against time.
- If this plot shows a rectangular band of scatter around the zero level, with no notable trends over time, this indicates that the specified model is adequate.
- See the residual plot for an $AR(1)$ modeling of the color property series.
- See the residual plot from an $AR(3)$ model fit for the square-root transformed hare data.
- See the residual plot from an $IMA(1,1)$ model fit for the logged oil price time series.

Q-Q Plots to Check Normality

- Determining whether the error terms are normally distributed in a time series model can be useful, since some inferences assume normal errors.
- A normal Q-Q plot of the residuals is a graphical check for normal errors.
- If the Q-Q plot resembles a straight line, then the assumption that the errors are normally distributed is reasonable.
- The Shapiro-Wilk test is a formal hypothesis test for normality.
- The null hypothesis is that the errors are normal; a small p-value would cause us to doubt the normality assumption.
- See the R examples on the course web page.

Checking Independence of Errors

- If the noise terms are truly white noise, they should be uncorrelated.
- However, the residuals from even a correctly specified model can have nonzero autocorrelations, especially for smaller lags.
- If the sample ACF of the residuals shows autocorrelations significantly different from zero, we may need to change the model.
- The naive bounds $\pm 2/\sqrt{n}$ can be used as a rough guide of significance; if sample autocorrelations stay well within these bounds, the autocorrelation can be assumed to be minimal (see R example).
- We should pay close attention to autocorrelations at lags $12, 24, \ldots$ for monthly data, and $4, 8, \ldots$ for quarterly data. Excessive autocorrelation at these lags can indicate we need to use a model that accounts for seasonality.

Ljung-Box Test for Serial Dependence

- The Ljung-Box test checks whether the entire set of residual correlations is larger than we would expect to see if the correct ARMA-type model was specified.
- In any $ARMA(p,q)$ model (which includes $AR(p)$ and $MA(q)$ models as special cases), the test statistic is

$$Q^* = n(n+2)\left(\frac{\hat{r}_1}{n-1} + \frac{\hat{r}_2}{n-2} + \cdots + \frac{\hat{r}_K}{n-K}\right).$$

- If Q^* is large relative to a $\chi^2_{K-(p+q)}$ distribution (i.e., if the test's p-value is very small), then we conclude that the model is not appropriate due to the large residual autocorrelations.
- The maximum lag K is chosen fairly large so that lags beyond K are negligible; it may be wise to perform the test for a range of values of K.
- The `tsdiag` function in R plots p-values of the Ljung-Box test across a series of values of K.
- See the R examples on the color property data.

Runs Test for Dependence of Errors

- In Chapter 3, we saw the runs test, which can assess whether a time series can be viewed as independent.
- The runs test can be applied to the residuals to check whether the errors are dependent.
- If the p-value of the runs test on the residuals is very small (say, less than 0.05), we can reject the hypothesis of independent errors.

Overfitting

- A reasonable strategy to check whether our proposed model is reasonable is to *overfit* with a slightly more general model.
- The proposed model should be a special case of the more general model.
- For example, if we believe an $AR(2)$ model is appropriate, we could overfit the data with an $AR(3)$ model, which includes one additional parameter, ϕ_3 in this case.
- The proposed $AR(2)$ would be appropriate if:
  1. The additional parameter (say, ϕ_3 in the above example) is not significantly different from zero in the overfit model, and
  2. the estimates of the other parameter(s) do not change much from the proposed model to the overfit model.

Overfitting Example: Color Property Data

- See the R example of overfitting the color property data with an $AR(2)$ model to check the adequacy of our proposed $AR(1)$ model.
- A different way of overfitting for the color property data would be to try an $ARMA(1,1)$ model to check the adequacy of our proposed $AR(1)$ model.
- Note that both the $AR(2)$ model and the $ARMA(1,1)$ model include the $AR(1)$ model as a special case.
- For the color property data, the evidence from each overfit model supports the original choice of an $AR(1)$ model.
- See the R code for examples of residual analysis and overfitting with other time series.

Parameter Redundancy (Lack of Identifiability)

- When generalizing a model, it is possible to get *parameter redundancy*.
- Page 187 of the text gives the example of an $ARMA(1,2)$ model that can also be represented as an $ARMA(2,3)$ model.
- But in the $ARMA(2,3)$ model, the parameters are not uniquely identifiable.
- For *any value* of a particular constant, the model holds, and so the parameters can have infinitely many sets of values that would yield an equally good model.
- Clearly this lack of identifiability is a problem in a parametric model, and so we would not want to use the $ARMA(2,3)$ model here.

Advice for the Overfitting Strategy

- Carefully specify the original model, using evidence such as the ACF, PACF, and/or EACF, and any practical knowledge about the data process you might have.
- If a simple model fits well, it is usually preferable to a more complicated model.
- When overfitting, you should not increase the orders of the AR component and the MA component simultaneously.
- Use the residual analysis to give you clues about how to extend the model.
- For example, if you fit an $MA(1)$ model and the residual analysis shows substantial lag-2 correlation in the residuals, then try overfitting with an $MA(2)$ model rather than with an $ARMA(1,1)$ model.