Test on 'Econometric Methods in Forecasting'

November 2006

1. A time series of 4 observations is given as $(1, 2, 3, 4)$.

   (a) Apply SES (*single exponential smoothing*) with the specification $\hat{x}_1 = x_1$, $\alpha = 0.5$. First determine $\hat{x}_2, \ldots, \hat{x}_4$. [In case you forgot the formula: $\hat{x}_t = \alpha x_t + (1 - \alpha)\hat{x}_{t-1}$]

   (b) Determine $\hat{x}_4(1)$ and $\hat{x}_4(2)$, i.e. forecasts for the unknown values x_5, x_6.

   (c) In consideration of the recognizable structure of the given time series, would you apply SES at all? Which alternative method would you suggest?

2. You assume that the data-generating process follows an ARMA model, in symbols
$$X_t = \varphi X_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2},$$
where observations $X_1, \ldots, X_n$ are available. You also assume $\mathrm{E}(\varepsilon_t | \varepsilon_{t-1}, \ldots) = 0$.

   (a) Evaluate $\mathrm{E}(X_{t+1} | X_1, \ldots, X_t)$. You know that this conditional expectation is routinely used as a forecast for X_{t+1}, i.e. $\hat{X}_t(1)$. You may assume that $\varepsilon_t, \varepsilon_{t-1}, \ldots$ (of course, not ε_{t+1}) and that all parameter values are known.

   (b) Assume you know the 'true' parameter values $\varphi = \theta_1 = \theta_2 = 0.5$. In the data series of example # 1, substitute $\varepsilon_0 = \varepsilon_1 = 0$ and determine $\hat{X}_3(1)$. [Hint: the result is a number. You are supposed to first determine ε_2, ε_3 and then to evaluate $\hat{X}_3(1)$]

   (c) Why do we not simply assume that ε_t is *white noise* ? Comment.

3. Someone wishes to evaluate single-step forecasts for the variable X based on several time-series models, in order to find the best forecasting model. Data are available for $t = 1, \ldots, 100$. The following strategies are considered.

   i All model parameters are estimated from the total sample, forecasts $\hat{X}_{10}(1)$, $\hat{X}_{11}(1)$ etc. are evaluated according to formulae such as the one in example # 2. Squared prediction errors are averaged for $\left(\hat{X}_{10} - X_{11}\right)^2, \ldots, \left(\hat{X}_{99}(1) - X_{100}\right)^2$, and the model with the smallest value is viewed as the best prediction model.

ii All model parameters are estimated from the sample portion $X_1, \ldots, X_{60}$. Then, these estimates are used to determine $\hat{X}_{60}(1), \ldots, \hat{X}_{99}(1)$. Squared errors are evaluated, averaged and compared.

iii Model parameters are estimated from samples $X_1, \ldots, X_p$. Then, predictions $\hat{X}_p(1)$ are determined, and $\left(\hat{X}_p(1) - X_{p+1}\right)^2$ is evaluated. The sample end p is varied over 60,…,99. Again, squared errors are averaged, and statistics are compared across all models.

Which of these strategies would be your favorite and why? Why are the other two strategies insufficient?