# Chapter 9: Forecasting

- One of the critical goals of time series analysis is to forecast (predict) the values of the time series at times in the future.
- When forecasting, we ideally should evaluate the precision of the forecast.
- We will consider examples of forecasts for
  1. deterministic trend models;
  2. ARMA- and ARIMA-type models;
  3. models containing deterministic trends and ARMA (or ARIMA) stochastic components.
- The methods we use here assume the model (including parameter values) is known exactly.
- This is not true in practice, but for large sample sizes, the parameter estimates should be close to the true parameter values.

# Minimum MSE Forecasting

- Assume we have observed the time series up to the present time, $t$, so that we have observed $Y_1, Y_2, \ldots, Y_t$.
- The goal is to forecast the value of $Y_{t+\ell}$, which is the value $\ell$ time units into the future.
- In this case, time $t$ is called the *forecast origin* and $\ell$ is called the *lead time* of the forecast.
- The forecast (predicted future value) itself is denoted $\hat{Y}_t(\ell)$.
- We will find the forecast formula that minimizes the mean square error (MSE) of the forecast, $E[(Y_{t+\ell} - \hat{Y}_t(\ell))^2]$, for a variety of models.
- The formula for our forecast $\hat{Y}_t(\ell)$ is the expected value of $Y_{t+\ell}$, *conditional* on the already-observed values of the series:

$$\hat{Y}_t(\ell) = E(Y_{t+\ell} | Y_1, \ldots, Y_t)$$

# Forecasting with a Deterministic Trend Model

- Consider the trend model $Y_t = \mu_t + X_t$, where $\mu_t$ is some deterministic trend and the stochastic component $X_t$ has mean zero.
- In particular, we assume $\{X_t\}$ is white noise with variance $\gamma_0$. Then

$$\begin{aligned}\hat{Y}_t(\ell) &= E(\mu_{t+\ell} + X_{t+\ell} | Y_1, Y_2, \ldots, Y_t) \\ &= E(\mu_{t+\ell} | Y_1, Y_2, \ldots, Y_t) + E(X_{t+\ell} | Y_1, Y_2, \ldots, Y_t) \\ &= E(\mu_{t+\ell}) + E(X_{t+\ell}) = \mu_{t+\ell},\end{aligned}$$

since $X_{t+\ell}$ has mean zero and is independent of the previously observed values $Y_1, Y_2, \ldots, Y_t$.

# Forecasting with a Linear Trend Model

- In the case in which we assume a linear trend, $\mu_t = \beta_0 + \beta_1 t$.
- So the forecast of the response at $\ell$ time units into the future is $\hat{Y}_t(\ell) = \beta_0 + \beta_1(t + \ell)$.
- This forecast assumes that the same linear trend holds in the future, which can be a dangerous assumption, since we don't have the (future) data (yet) to justify it.

# Forecasting with Other Trend Models

- For a quadratic trend, where $\mu_t = \beta_0 + \beta_1 t + \beta_2 t^2$, the forecast is $\hat{Y}_t(\ell) = \beta_0 + \beta_1(t+\ell) + \beta_2(t+\ell)^2$.
- With higher-order polynomial trends, extrapolating into the future becomes even more risky.
- For periodic seasonal means models in which $\mu_t = \mu_{t+12}$, the forecast is $\hat{Y}_t(\ell) = \mu_{t+12+\ell} = \hat{Y}_t(\ell + 12)$.
- So for such models, the forecast at a particular time is the same as the forecast at the time 12 months later.
- See the examples of forecasts on real data sets on the course web page.

# Forecast Error and Forecast Error Variance

- The *forecast error* is denoted by $e_t(\ell)$:

$$\begin{aligned} e_t(\ell) &= Y_{t+\ell} - \hat{Y}_t(\ell) \\ &= \mu_{t+\ell} + X_{t+\ell} - \mu_{t+\ell} = X_{t+\ell}, \end{aligned}$$

  so that $E[e_t(\ell)] = E[X_{t+\ell}] = 0$.
- Thus the forecast is *unbiased*.
- And the *forecast error variance* is $var[e_t(\ell)] = var[X_{t+\ell}] = \gamma_0$, which does not depend on the lead time $\ell$.

# Forecasting in $AR(1)$ Models

- Consider the $AR(1)$ process with a nonzero mean $\mu$:

$$Y_t - \mu = \phi(Y_{t-1} - \mu) + e_t.$$

- Suppose we want to forecast the process 1 time unit into the future. Note that

$$Y_{t+1} - \mu = \phi(Y_t - \mu) + e_{t+1}.$$

- Taking the conditional expected value (given $Y_1, Y_2, \ldots, Y_t$) of both sides, we have:

$$\begin{aligned}\hat{Y}_t(1) - \mu &= \phi[E(Y_t | Y_1, Y_2, \ldots, Y_t) - \mu] + E(e_{t+1} | Y_1, Y_2, \ldots, Y_t) \\ &= \phi[Y_t - \mu] + E(e_{t+1}) = \phi[Y_t - \mu].\end{aligned}$$

since $e_{t+1}$ is independent of $Y_1, Y_2, \ldots, Y_t$ and has mean zero.

# Forecasting and the Difference Equation Form

- So $\hat{Y}_t(1) = \mu + \phi(Y_t - \mu)$.
- That is, the forecast for the next value is the process mean, plus some fraction of the current deviation from the process mean.
- If we forecast not just 1 time unit but $\ell$ time units into the future, we have

$$\hat{Y}_t(\ell) = \mu + \phi[\hat{Y}_t(\ell - 1) - \mu] \text{ for } \ell \geq 1.$$

- So any forecast can be found recursively: We can find $\hat{Y}_t(1)$, which we can then use to find $\hat{Y}_t(2)$, etc.
- This recursive formula is called the *difference equation* form of the forecasts.

# A General Formula for Forecasts in $AR(1)$ Models

- Note that we can solve for a general formula for a forecast with a lead time $\ell$ in an $AR(1)$ process:

$$\begin{aligned}
\hat{Y}_t(\ell) &= \phi[\hat{Y}_t(\ell-1)-\mu]+\mu \\
&= \phi[\{\phi[\hat{Y}_t(\ell-2)-\mu]\}+\mu-\mu]+\mu \\
&= \phi[\{\phi[\hat{Y}_t(\ell-2)-\mu]\}]+\mu \\
&\vdots \\
&= \phi^{\ell-1}[\hat{Y}_t(1)-\mu]+\mu \\
&= \phi^{\ell-1}[\mu+\phi(Y_t-\mu)-\mu]+\mu
\end{aligned}$$

  which implies that $\hat{Y}_t(\ell)=\mu+\phi^{\ell}(Y_t-\mu)$.
- So the fraction of the current deviation from the process mean that is added to $\mu$ becomes closer to zero as the lead time gets larger.

# Forecasting with the Color Property Example

- Recall that we used a $AR(1)$ model for the color property time series.
- Via ML, we estimated $\phi$ and $\mu$ to be 0.5705 and 74.3293, respectively.
- For the purpose of the forecast, we will take these to be the true parameter values (though they really are not).
- The last observed value, $Y_t$, of this color property series was 67.
- So forecasting 1 time unit into the future yields $\hat{Y}_t(1) = 74.3293 + 0.5705(67 - 74.3293) = 70.14793$.

# Forecasting with the Color Property Example (continued)

- To forecast, say, 5 time units into the future, we can continue recursively, or just use the general formula to obtain: $\hat{Y}_t(5) = 74.3293 + 0.5705^5(67 - 74.3293) = 73.88636.$
- Note that forecasting 20 time units into the future yields $\hat{Y}_t(20) = 74.3293 + 0.5705^{20}(67 - 74.3293) = 74.3292.$
- We see that for a large lead time, the forecast nearly equals $\mu$.
- In general, for all stationary ARMA models, $\hat{Y}_t(\ell) \approx \mu$ for large $\ell$.

# One-step-ahead Forecast Error

- The *one-step-ahead forecast error* $e_t(1)$ is the difference between the actual value of the process one time unit into the future and the predicted value one time unit ahead.
- For the $AR(1)$ model, this is $e_t(1) = Y_{t+1} - \hat{Y}_t(1) = [\phi(Y_t - \mu) + \mu + e_{t+1}] - [\phi(Y_t - \mu) + \mu] = e_{t+1}$.
- So the one-step-ahead forecast error is simply a white-noise observation, and it is independent of $Y_1, Y_2, \ldots, Y_t$.
- And $var[e_t(1)] = \sigma_e^2$.

# Forecast Error for General Lead Time

- The *forecast error* for a general lead time, $\ell$, $e_t(\ell)$, is the difference between the actual value of the process $\ell$ time units into the future and the predicted value $\ell$ time units ahead.
- For any general linear process, it can be shown that

$$e_t(\ell) = e_{t+\ell} + \psi_1 e_{t+\ell-1} + \psi_2 e_{t+\ell-2} + \cdots + \psi_{\ell-1} e_{t+1}$$

- Clearly, $E[e_t(\ell)] = 0$, so the forecasts are unbiased.
- And $var[e_t(\ell)] = \sigma_e^2(1 + \psi_1^2 + \psi_2^2 + \cdots + \psi_{\ell-1}^2)$.
- These results hold for all ARIMA models.

# Forecast Error for General Lead Time in $AR(1)$ Models

- For an $AR(1)$ process, the forecast error for a general lead time is

$$e_t(\ell) = e_{t+\ell} + \phi e_{t+\ell-1} + \phi^2 e_{t+\ell-2} + \cdots + \phi^{\ell-1} e_{t+1}$$

- And $var[e_t(\ell)] = \sigma_e^2 \left[ \frac{1-\phi^{2\ell}}{1-\phi^2} \right]$.
- So for long lead times, $var[e_t(\ell)] \approx \frac{\sigma_e^2}{1-\phi^2}$ for large $\ell$.
- And since this right hand side is the variance formula for an $AR(1)$ process, note that $var[e_t(\ell)] \approx var(Y_t) = \gamma_0$ for large $\ell$.
- This last result holds for all stationary ARMA models.

# Forecasting with an $MA(1)$ Model

- Consider now an $MA(1)$ model with a nonzero mean, $Y_t = \mu + e_t - \theta e_{t-1}$.
- Replacing $t$ by $t + 1$ and taking conditional expectations, we have
$$\hat{Y}_t(1) = \mu - \theta E(e_t | Y_1, Y_2, \dots, Y_t).$$
- If the model is invertible, then $E(e_t | Y_1, Y_2, \dots, Y_t) = e_t$ (at least approximately, since we condition on $Y_1, Y_2, \dots, Y_t$ rather than on the infinite history $\dots, Y_0, Y_1, Y_2, \dots, Y_t$).
- If the model is not invertible, then $E(e_t | Y_1, Y_2, \dots, Y_t) \neq e_t$ (not even approximately).
- For an invertible $MA(1)$ model, the one-step-ahead forecast is $\hat{Y}_t(1) = \mu - \theta e_t$.

# Forecast Error for $MA(1)$ Model

- Again, the one-step-ahead forecast error is $e_t(1) = Y_{t+1} - \hat{Y}_t(1) = [\mu + e_{t+1} - \theta e_t] - [\mu - \theta e_t] = e_{t+1}$.
- For longer lead time, where $\ell > 1$,

$$\hat{Y}_t(\ell) = \mu + E(e_{t+\ell} | Y_1, Y_2, \ldots, Y_t) - \theta E(e_{t+\ell-1} | Y_1, Y_2, \ldots, Y_t)$$

- But for $\ell > 1$, both $e_{t+\ell}$ and $e_{t+\ell-1}$ are independent of $Y_1, Y_2, \ldots, Y_t$, so these conditional expected values are both zero.
- Therefore, in an invertible $MA(1)$ model, $\hat{Y}_t(\ell) = \mu$ for $\ell > 1$.

# Forecasting with the Random Walk with Drift

- Now we consider forecasting with a nonstationary ARIMA process.
- Specifically, consider the *random walk with drift* model, where $Y_t = Y_{t-1} + \theta_0 + e_t$.
- This is basically an $ARIMA(0, 1, 0)$ model with an extra constant term.
- The forecast one step ahead is

$$\begin{aligned}\hat{Y}_t(1) &= E(Y_t | Y_1, Y_2, \ldots, Y_t) + \theta_0 + E(e_{t+1} | Y_1, Y_2, \ldots, Y_t) \\ &= Y_t + \theta_0\end{aligned}$$

# Forecasting with the Random Walk with Drift with General Lead Time

- For $\ell > 1$, $\hat{Y}_t(\ell) = \hat{Y}_t(\ell - 1) + \theta_0$.
- So by iterating backward, we see that $\hat{Y}_t(\ell) = Y_t + \theta_0 \ell$ for $\ell \geq 1$.
- The forecast, as a function of the lead time $\ell$, is a straight line with slope $\theta_0$.
- With nonstationary series, the presence of the constant term has a major effect on the forecast, so it is important to determine whether the constant term is truly needed (we could check whether it is significantly different from zero).

# Forecast Error with the Random Walk with Drift

- For the random walk with drift model, the one-step-ahead forecast error is again $e_t(1) = Y_{t+1} - \hat{Y}_t(1) = e_{t+1}$.
- But the forecast error $\ell$ steps ahead can be shown to be $e_t(\ell) = e_{t+1} + e_{t+2} + \cdots + e_{t+\ell}$.
- So $var[e_t(\ell)] = \ell\sigma_e^2$.
- In this *nonstationary* model, the variance of the forecast error continues to increase *without bound* as the lead time gets larger.
- This phenomenon will happen with all nonstationary ARIMA models.
- On the other hand, with stationary models, the variance of the forecast error increases as the lead time gets larger, but there is a limit to the increase.
- And with deterministic trend models, the variance of the forecast error is constant as the lead time gets larger.

# Forecasting with the $ARMA(p, q)$ Model

- The general difference equation form for forecasts in the $ARMA(p, q)$ model is somewhat complicated:

$$\begin{aligned}\hat{Y}_t(\ell) = {} & \phi_1 \hat{Y}_t(\ell - 1) + \phi_2 \hat{Y}_t(\ell - 2) + \cdots + \phi_p \hat{Y}_t(\ell - p) + \theta_0 \\ & - \theta_1 e_{t+\ell-1} I[\ell \leq 1] - \theta_2 e_{t+\ell-2} I[\ell \leq 2] \\ & - \cdots - \theta_q e_{t+\ell-2} I[\ell \leq q]\end{aligned}$$

  where the indicator $I[\cdot]$ equals 1 if the condition in the brackets is true, and 0 otherwise.
- For example, with an $ARMA(1, 1)$ model, $\hat{Y}_t(1) = \phi Y_t + \theta_0 - \theta e_t$, and $\hat{Y}_t(2) = \phi \hat{Y}_t(1) + \theta_0$, and in general, $\hat{Y}_t(\ell) = \phi \hat{Y}_t(\ell - 1) + \theta_0$ for $\ell \geq 2$.
- With an $ARMA(1, 1)$ model, an explicit general formula for a forecast $\ell$ time units ahead, in terms of $\mu = E(Y_t)$, is

$$\hat{Y}_t(\ell) = \mu + \phi^\ell (Y_t - \mu) - \phi^{\ell-1} \theta e_t \text{ for } \ell \geq 1.$$

# More On Forecasting with the $ARMA(p, q)$ Model

- For lead time $\ell = 1, 2, \ldots, q$, the noise terms appear in the formulas for the forecasts.
- When calculating the forecast, noise terms for past times $(\leq t)$ are replaced by the corresponding residuals, and future noise terms are replaced by zero.
- For longer lead times (i.e., $\ell > q$) the noise terms disappear and only the autoregressive component (and the constant term) of the model affects the forecast.
- For $\ell > q$, the difference equation formula for the $ARMA(p, q)$ model reduces to $\hat{Y}_t(\ell) = \phi_1 \hat{Y}_t(\ell - 1) + \phi_2 \hat{Y}_t(\ell - 2) + \cdots + \phi_p \hat{Y}_t(\ell - p) + \theta_0$.

# Forecasting with the $ARMA(p, q)$ Model as Lead Times Increase

- Since we have shown that $\theta_0 = \mu(1 - \phi_1 - \phi_2 - \cdots - \phi_p)$, this can be rewritten as

$$\hat{Y}_t(\ell) - \mu = \phi_1[\hat{Y}_t(\ell - 1) - \mu] + \phi_2[\hat{Y}_t(\ell - 2) - \mu] + \cdots + \phi_p[\hat{Y}_t(\ell - p) - \mu] \text{ for } \ell \geq q.$$

- For a stationary ARMA model, $\hat{Y}_t(\ell) - \mu$ will decay toward zero as the lead time $\ell$ increases, and thus for long lead times, the forecast will approximately equal the process mean $\mu$.
- This is sensible because for stationary models, the dependence grows weaker as the time between observations increases, and $\mu$ would be the natural best forecast to use if there were *no* dependence over time.

# Forecasting with Nonstationary Models

- We have seen one example of forecasting with nonstationary models (the random walk with drift).
- For an $ARIMA(1,1,1)$ model,

$$\begin{aligned}
\hat{Y}_t(1) &= (1+\phi)Y_t - \phi Y_{t-1} + \theta_0 - \theta e_t \\
\hat{Y}_t(2) &= (1+\phi)\hat{Y}_t(1) - \phi Y_t + \theta_0 \\
&\vdots \\
\hat{Y}_t(\ell) &= (1+\phi)\hat{Y}_t(\ell-1) - \phi \hat{Y}_t(\ell-2) + \theta_0
\end{aligned}$$

- These forecasts are unbiased, i.e., $E[e_t(\ell)] = 0$ for any $\ell \geq 1$.

# Forecast Error Variance with Nonstationary Models

- But the variance of the forecast error is

$$var[e_t(\ell)] = \sigma_e^2 \sum_{j=0}^{\ell-1} \psi_j^2 \text{ for } \ell \geq 1.$$

- For a nonstationary series, these $\psi_j$ weights do not decay to zero as $j$ increases.
- So the forecast error variance increases without bound as the lead time $\ell$ increases.
- Lesson: With nonstationary series, when we forecast far into the future, we have a lot of uncertainty about the forecast.