1. The Box-Cox transformation is given by

\[ Z_t = \begin{cases} \frac{(Z_t^* + m)^{\gamma} - 1}{\gamma} & \gamma \neq 0 \\ \log(Z_t^* + m) & \gamma = 0 \end{cases} \]

Usually we take \( m = 0 \) but there are some circumstances where we would use another value of \( m \). Give an example if a situation where we would want to use \( m > 0 \).

If a time series contains some values that are close to 0, the logarithms of those values will be close to \(-\infty\). Adding a constant \( m \) to all of the values of the time series will weaken the effect of the transformation.

2. If you had a time series process consisting of counting the number of events that are relatively rare (e.g., the number of serious automobile accidents in Ames in a month), what transformation might you expect to provide more constant variance than the original data. Explain why.

As discussed in class a Poisson distribution is often used to model count data, but the mean and variance of this distribution are the same implying nonconstant variance if the mean changes. Using a square root transformation, however, will make the variance approximately constant.

3. Consider the ARMA(1,1) model

\[ (1 - \phi_1 B)Z_t = (1 - \theta_1 B)a_t \]

As we know, this model can be expressed as

\[ Z_t = \psi_1 a_{t-1} + \psi_2 a_{t-2} + \psi_3 a_{t-3} + \cdots + a_t \]

(a) Derive expressions for \( \psi_1 \) and \( \psi_2 \) for this model.

\[ Z_t = (1 - \phi_1 B)^{-1} (1 - \theta_1 B) a_t \]

Then using the geometric series,

\[ = (1 + \phi_1 B + \phi_1^2 B^2 + \cdots) (1 - \theta_1 B) a_t \]

\[ = (1 + \phi_1 B + \phi_1^2 B^2 + \cdots) (1 - \theta_1 B - \theta_1 \phi_1 B^2 - \theta_1 \phi_1^2 B^3 + \cdots) a_t \]

\[ = (\phi_1 - \theta_1) a_{t-1} + \phi_1 (\phi_1 - \theta_1) a_{t-2} + \phi_1^2 (\phi_1 - \theta_1) a_{t-3} + \cdots + a_t \]

\[ \psi_1 = \phi_1 - \theta_1, \quad \psi_2 = \phi_1 (\phi_1 - \theta_1) \]

(b) Derive \( Cov(Z_t, a_{t-2}) \) for this model.

\[ Cov(Z_t, a_{t-2}) = E(Z_t a_{t-2}) = \psi_2 = \phi_1 (\phi_1 - \theta_1) \]

\[ = E(\psi_1 a_{t-1} a_{t-2} + \psi_2 a_{t-2} a_{t-2} + \psi_3 a_{t-3} a_{t-2} + \cdots) \]

\[ = E(\psi_2 a_{t-2}^2) = \psi_2 \sigma_a^2 = \phi_1 (\phi_1 - \theta_1) \sigma_a^2 \]
4. A time series $Z_t$ can be described by a random walk $Z_t = Z_{t-1} + a_t$. Assume that a realization $Z_1, Z_2, \ldots, Z_{100}$ is available to compute forecasts and that a forecast is needed for $Z_{103}$, using $Z_{100}$ as the forecast origin.

(a) Give an expression for $Z_{103}$, based on this random walk model.

$$Z_{103} = Z_{100} + a_{101} + a_{102} + a_{103}$$

(b) Give an expression that can be used to compute $\hat{Z}_{100}(3)$, the forecast for $Z_{103}$.

$$\hat{Z}_{100}(3) = Z_{100}$$

(c) Derive $e_{100}(3)$, the forecast error in $\hat{Z}_{100}(3)$.

$$e_{100}(3) = Z_{103} - \hat{Z}_{100}(3) = a_{101} + a_{102} + a_{103}$$

(d) Give an expression for the variance of $e_{100}(3)$.

$$\text{Var}(e_{100}(3)) = \text{Var}(a_{101} + a_{102} + a_{103}) = 3 \sigma_a^2$$

(e) Give an expression for an approximate 95% prediction interval for $Z_{103}$.

$$Z_{100} \pm 1.96 \sqrt{3 \sigma_a^2}$$
5. Which diagnostic is most useful for detecting a violation of the assumption that the $a_t$ values are independent?

\[ ACF \text{ of the residuals} \]

6. The AR(1) model can be expressed by

\[ Z_t = \theta_0 + \phi_1 Z_{t-1} + a_t, \quad a_t \sim \text{nid}(0, \sigma_a^2). \]

\[ \hat{Z}_t = \phi_1 \hat{Z}_{t-1} + \alpha_t \]

(a) Derive an expression for the variance $\sigma_Z^2 = \text{Var}(Z)$ for this model.

\[ \gamma_0 = \text{Var}(\hat{Z}_t) = E(\hat{Z}_t^2) = E(\phi_1 \hat{Z}_{t-1} \hat{Z}_t + \hat{Z}_t a_t) = \phi_1 \gamma_1 + \sigma_a^2 \]

\[ \frac{\gamma_0}{\gamma_0} = 1 = \phi_1 \rho_1 + \frac{\sigma_a^2}{\gamma_0} \Rightarrow 1 - \phi_1 \rho_1 = \frac{\sigma_a^2}{\gamma_0} \Rightarrow \frac{\gamma_0}{\sigma_a^2} = 1/(1 - \phi_1 \rho_1) \]

\[ \Rightarrow \gamma_0 = \frac{\sigma_a^2}{1 - \phi_1 \rho_1} \]

(b) Derive expressions for the autocorrelation function values $\rho_1$ and $\rho_2$ for this model.

\[ \rho_1 = \frac{\gamma_1}{\gamma_0} = \phi_1 \]

\[ \rho_2 = \phi_1 \rho_1 = \phi_1^2 \]

(c) Explain how to compute (i.e., give a formula for) the first two true PACF values $\phi_{11}$ and $\phi_{22}$ for this model.

\[ \phi_{11} = \rho_1 \]

\[ \phi_{22} = \frac{\rho_2 - \phi_1 \rho_1}{1 - \phi_1^2} = \frac{\phi_1^2 - \phi_1^2}{1 - \phi_1^2} = 0 \]

7. The AIC has been suggested as a useful tool for choosing among time series models.

(a) Why is AIC a better criterion than choosing the model with the largest likelihood?

If you add a new term to a model, the likelihood will never decrease.

AIC adds a penalty corresponding to the number of parameters being estimated.

(b) Explain why AIC should not be used exclusively in making such a choice. Be as specific as possible in your answer.

It is important to also consider and use knowledge of the data-generating process and diagnostics when choosing a model.
8. Suppose that the model for $Z_t$ is $(1 - B)Z_t = \theta_0 + a_t$ and that the working series is defined by $W_t = (1 - B)^2 Z_t$.

\[ Z_t = (1-B)^{-1} (\theta_0 + a_t) \]

\[ W_t = (1-B)^2 (1-B)^{-1} (\theta_0 + a_t) \]

\[ = (1-B) (\theta_0 + a_t) \]

\[ = (\theta_0 + a_t) - (\theta_0 + a_{t-1}) = a_t - a_{t-1} = (1-B)a_t \]

(a) Derive an expression for the model for $W_t$ (note that $Z_t$ should not appear in this expression).

(b) Derive the mean of $W_t$.

\[ E(W_t) = E(a_t) - E(a_{t-1}) = 0 - 0 = 0 \]

(c) Derive the variance of $W_t$.

\[ \text{Var}(W_t) = E(W_t^2) = E(a_t^2 - 2a_t a_{t-1} + a_{t-1}^2) \]

\[ = E(a_t^2) - 2E(a_t a_{t-1}) + E(a_{t-1}^2) \]

\[ = \sigma_a^2 + 0 + \sigma_a^2 = 2\sigma_a^2 \]

(d) Is $W_t$ stationary? Why or why not?

$W_t$ is stationary because its mean, variance, and ACF do not depend on $t$.

(e) Is $W_t$ invertible? Why or why not?

$W_t$ is not invertible because it is a MA(1) model, with a root on the unit circle.