Estimation and Forecasting

Exercise 1 (Yule-Walker estimation)
Suppose we observe $n = 250$ data points generated by an AR(2) process,

$$x_t = 1.5x_{t-1} - 0.75x_{t-2} + \varepsilon_t, \tag{1}$$

that is, the datapoints were simulated with this model. Now suppose that we are only given these data without knowing the true model specified above. For these data, we can compute the following quantities: $\hat{\gamma}(0) = 8.1281$, $\hat{\rho}(1) = 0.8645$, $\hat{\rho}(2) = 0.5769$.

(a) Compute the parameter estimates for α_1 and α_2 using the Yule-Walker method.

The **Yule-Walker estimators** are given by

$$\begin{aligned} \hat{\alpha} &= \hat{\Gamma}_p^{-1}\hat{\gamma}_p, \\ \hat{\sigma}^2 &= \hat{\gamma}(0) - \hat{\alpha}\hat{\gamma}_p \\ &= \hat{\gamma}(0) - \hat{\gamma}_p'\hat{\Gamma}_p^{-1}\hat{\gamma}_p. \end{aligned}$$

More conveniently, we equivalently write the estimator in terms of the sample ACF (factoring out $\hat{\gamma}(0)$), namely

$$\begin{aligned} \hat{\alpha} &= \hat{R}_p^{-1}\hat{\rho}_p, \\ \hat{\sigma}^2 &= \hat{\gamma}(0)\left[1 - \hat{\rho}_p'\hat{R}_p^{-1}\hat{\rho}_p\right]. \end{aligned}$$

For our data at hand, we get

$$\begin{aligned} \hat{\alpha} = \begin{pmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \end{pmatrix} &= \begin{pmatrix} 1 & \hat{\rho}(1) \\ \hat{\rho}(1) & 1 \end{pmatrix}^{-1} \begin{pmatrix} \hat{\rho}(1) \\ \hat{\rho}(2) \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0.8645 \\ 0.8645 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 0.8645 \\ 0.5769 \end{pmatrix} \\ &= \begin{pmatrix} 3.9582 & -3.4219 \\ -3.4219 & 3.9582 \end{pmatrix} \begin{pmatrix} 0.8645 \\ 0.5769 \end{pmatrix} = \begin{pmatrix} 1.4478 \\ -0.6747 \end{pmatrix}. \end{aligned}$$

Exact solution:

$$\hat{\alpha} = \frac{1}{1 - \hat{\rho}(2)^2} \begin{pmatrix} \hat{\rho}(1)(1 - \hat{\rho}(2)) \\ \hat{\rho}(2) - \hat{\rho}(2)^2 \end{pmatrix} = \begin{pmatrix} 1.4478 \\ -0.6747 \end{pmatrix}. \tag{2}$$

(b) What is the estimated variance of the innovation process $\{\varepsilon_t\}_t$?

$$\begin{aligned}\hat\sigma^2 &= \hat\gamma(0)[1-\hat\rho(1)\hat\alpha_1-\hat\rho(2)\hat\alpha_2]\\ &= 8.1281\left[1-\begin{pmatrix}0.8645 & 0.5769\end{pmatrix}\begin{pmatrix}1.4478\\ -0.6747\end{pmatrix}\right]\\ &= 8.1281[1-0.8624] = 1.1184.\end{aligned}$$

Exact solution:

$$\hat\sigma^2 = \hat\gamma(0)\frac{1-2\hat\rho(1)^2(1-\hat\rho(2))-\hat\rho(2)^2}{1-\hat\rho(1)^2} = 1.1186. \tag{3}$$

(c) Compute the estimated variance-covariance matrix of the parameter vector $\hat\alpha = (\hat\alpha_1, \hat\alpha_2)'$

Hint: The asymptotic ($n \to \infty$) behavior of the Yule-Walker estimators in the case of causal AR(p) processes is as follows:

$$\sqrt{n}(\hat\alpha - \alpha) \xrightarrow{d} N(0, \sigma^2\Gamma_p^{-1}), \qquad \hat\sigma^2 \xrightarrow{p} \sigma^2. \tag{4}$$

Use this to compute the covariance matrix with the estimates you obtained from the previous tasks.

Therefore, for our data at hand we obtain

$$\begin{aligned}\frac{1}{n}\frac{\hat\sigma^2}{\hat\gamma(0)}\hat R_p^{-1} &= \frac{1}{250}\frac{1.1184}{8.1281}\begin{pmatrix}1 & 0.8645\\ 0.8645 & 1\end{pmatrix}^{-1}\\ &= \frac{1}{250}\frac{1.1184}{8.1281}\begin{pmatrix}3.9582 & -3.4219\\ -3.4219 & 3.9582\end{pmatrix}\\ &= \begin{pmatrix}0.0022 & -0.0019\\ -0.0019 & 0.0022\end{pmatrix}\end{aligned}$$

as estimate for the variance-covariance matrix of the estimated parameters.

The exact solution is the same:

$$\frac{1}{n}\frac{\hat\sigma^2}{\hat\gamma(0)}\hat R_p^{-1} = \begin{pmatrix}0.0022 & -0.0019\\ -0.0019 & 0.0022\end{pmatrix} \tag{5}$$

(d) Use the last result to construct an approximate 95% confidence interval for $\hat\alpha_2$.

$$\hat\alpha_2 \pm 1.96\sqrt{\mathrm{Var}(\hat\alpha_2)} = -0.6747 \pm 1.96\cdot\sqrt{0.0022} = [-0.7666, -0.5828]. \tag{6}$$

$$\hat\alpha_2 \pm \alpha^{-1}(0.025)\sqrt{\mathrm{Var}(\hat\alpha_2)} = [-0.7662, -0.5832]. \tag{7}$$

Exercise 2 (ML estimation)
Consider the AR(1) process

$$x_t = \alpha_1 x_{t-1} + \varepsilon_t, \quad \varepsilon_t \overset{iid}{\sim} GWN(0,\sigma^2), \tag{8}$$

with $|\alpha_1| < 1$.

(a) State the unconditional density of x_1. *Hint: The process is causal.*

From the MA(∞)-representation (exists because the process is causal)

$$x_1 = \sum_{j=0}^{\infty} \alpha_1^j \varepsilon_{1-j} \tag{9}$$

it follows that x_1 is normally distributed with zero mean and variance $\frac{\sigma^2}{1-\alpha_1^2}$. Therefore, the density is given by

$$f_{x_1}(x_1; \alpha_1, \sigma^2) = \frac{1}{\sqrt{2\pi}\sqrt{\sigma^2/(1-\alpha_1^2)}} \exp\left\{ -\frac{1}{2} \frac{x_1^2}{\sigma^2/(1-\alpha_1^2)} \right\}. \tag{10}$$

(b) Determine the conditional density $f_{x_t|x_{t-1}}(x_t|x_{t-1})$.

x_t and x_{t-1} are normally distributed with zero mean and variance-covariance matrix

$$\Sigma = \frac{\sigma^2}{1-\alpha_1^2} \begin{pmatrix} 1 & \alpha_1 \\ \alpha_1 & 1 \end{pmatrix}. \tag{11}$$

Therefore, the conditional distribution of $x_t|x_{t-1} = \tilde{x}_{t-1}$ is also a normal distribution with mean $\alpha_1 x_{t-1}$ and variance σ^2. This can be either obtained by directly computing the conditional mean and variance from the autoregressive equation

$$\begin{aligned} E(x_t|x_{t-1} = \tilde{x}_{t-1}) &= E(\alpha_1 x_{t-1} + \varepsilon_t | x_{t-1} = \tilde{x}_{t-1}) = \alpha_1 \tilde{x}_{t-1} \\ \mathrm{Var}(x_t|x_{t-1} = \tilde{x}_{t-1}) &= \mathrm{Var}(\alpha_1 x_{t-1} + \varepsilon_t | x_{t-1} = \tilde{x}_{t-1}) = \mathrm{Var}(\varepsilon_t) = \sigma^2 \end{aligned}$$

or by using the general result for 2-dimensional normal distributions

$$y_1|y_2 = y_2 \sim \mathrm{N}\left(\mu_1 + \frac{\sigma_1}{\sigma_2}\rho(y_2 - \mu_2), (1-\rho^2)\sigma_1^2 \right). \tag{12}$$

For the conditional density we get:

$$f_{x_t|x_{t-1}}(x_t|x_{t-1}; \alpha_1, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{1}{2} \frac{(x_t - \alpha_1 x_{t-1})^2}{\sigma^2} \right\}. \tag{13}$$

(c) The process exhibits the so-called *Markov property*. Explain.

The conditional distribution of x_t given the entire past up to x_1 (or even the infinite past) depends only on x_{t-1}, i.e., all relevant information about the past is already included in x_{t-1}.

The future x_{t+1} is – if we know the present x_t – conditionally independent of the past $x_{t-1}, x_{t-2}, \ldots$, which is exactly the Markov property.

(d) Suppose you observe realizations $\tilde{x} = (\tilde{x}_1, \ldots, \tilde{x}_T)'$ of the process. Write down the likelihood and log-likelihood (with respect to α_1, σ^2 and $\tilde{x}$).

Using the Markov property, the likelihood function can be written as a product of the unconditional density of x_1 and the conditional densities of $x_2, \ldots, x_T$:

$$\begin{aligned} L(\tilde{x}; \alpha_1, \sigma^2) &= f_{x_1}(\tilde{x}_1, \alpha_1, \sigma^2) \prod_{t=2}^{T} f_{x_t|x_{t-1}}(\tilde{x}_t | x_{t-1}, \alpha_1, \sigma^2) \\ &= \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^T \left(\frac{1}{\sqrt{1/(1-\alpha_1^2)}} \right) \exp\left(-\left(\frac{\tilde{x}_1^2(1-\alpha_1^2)}{2\sigma^2} \right) - \sum_{t=2}^{T} \frac{(\tilde{x}_t - \alpha_1 \tilde{x}_{t-1})^2}{2\sigma^2} \right). \end{aligned}$$

For the log-likelihood it holds

$$
\begin{aligned}
l(\tilde{x};\alpha_1,\sigma^2) &= \ln(f_{x_1}(\tilde{x}_1,\alpha_1,\sigma^2)) + \sum_{t=2}^{T} \ln(f_{x_t|x_{t-1}}(\tilde{x}_t|\tilde{x}_{t-1},\alpha_1,\sigma^2)) \\
&= -\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln(\sigma^2) + \frac{1}{2}\ln(1-\alpha_1^2) - \left(\frac{\tilde{x}_1^2(1-\alpha_1^2)}{2\sigma^2}\right) - \sum_{t=2}^{T}\frac{(\tilde{x}_t-\alpha_1\tilde{x}_{t-1})^2}{2\sigma^2}.
\end{aligned}
$$

Forecasting

For the following exercises, we have to make some assumptions and introduce some notation.

- Throughout, we assume x_t is a causal and invertible ARMA(p,q) process, that is $\alpha(L)x_t = \beta(L)\varepsilon_t$.
- The minimum mean square error predictor based on the given data $\{x_n, x_{n-1}, ..., x_1\}$, is denoted by
$$x^n_{n+m} = \mathbb{E}[x_{n+m}|x_n, ..., x_1]$$
- If we assume to have knowledge of the complete history of the process,$\{x_n, x_{n-1}, ...\}$, we denote the predictor based on the infinite past by
$$\tilde{x}^n_{n+m} = \mathbb{E}[x_{n+m}|x_n, x_{n-1}, ...]$$
- Henceforth we confine attention to *linear* prediction.
Given data $x_1, ..., x_n$, the best linear predictor,
$$x^n_{n+m} = \alpha_0 + \sum_{k=1}^{n} \alpha_k x_k$$
of x_{n+m}, for m ≥ 1, is found by solving
$$\mathbb{E}[(x_{n+m} - x^n_{n+m})x_k] = 0, \qquad k = 0, 1, 2, ...n$$
where $x_0 = 1$

The above equations are called the *prediction equations*.

Exercise 3 (Forecasting MA(1) processes)
Consider the MA(1) process
$$x_t = \varepsilon_t + \beta_1\varepsilon_{t-1}$$
where $\{\varepsilon\}_t$ is white noise with variance σ^2.

(a) Derive the minimum mean square error one-step forecast based on the infinite past, and determine the mean square error of this forecast.

(b) Let $\tilde{x}^n_{n+1}$ be the truncated one-step-ahead forecast, i.e. the forecast based on the last n observations. (In principle the entire past of the process is known.) Show that
$$\mathbb{E}[(x_{n+1} - \tilde{x}^n_{n+1})^2] = \sigma^2(1 + \beta_1^{2+2n})$$
Compare the result with (a), and indicate how well the finite approximation works in this case.

(a) Write the model
$$x_t = \varepsilon_t + \beta_1\varepsilon_{t-1}$$

in its invertible form

$$
\begin{aligned}
\varepsilon_t &= -\beta_1 \varepsilon_{t-1} + x_t \\
&= -\beta_1(-\beta_1 \varepsilon_{t-2} + x_{t-1}) + x_t \\
&\vdots \\
&= \sum_{j=0}^{\infty} (-\beta_1)^j x_{t-j} \\
&= x_t + \sum_{j=1}^{\infty} (-\beta_1)^j x_{t-j}.
\end{aligned}
$$

Rearranging yields

$$
x_t = -\sum_{j=1}^{\infty} (-\beta_1)^j x_{t-j} + \varepsilon_t.
$$

One period later we have

$$
x_{t+1} = -\sum_{j=1}^{\infty} (-\beta_1)^j x_{t-j+1} + \varepsilon_{t+1}.
$$

Now replace the index t by n, implying that we are looking for the one-step-ahead forecast while knowing the past up to time n, and take conditional expectations, yielding the desired forecast

$$
\begin{aligned}
\tilde{x}_{n+1} &= E[x_{n+1} | x_n, x_{n-1}, \ldots] \\
&= -\sum_{j=1}^{\infty} (-\beta_1)^j x_{n-j+1} + \underbrace{E(\epsilon_{t+1})}_{=0}.
\end{aligned}
$$

For the MSE we obtain

$$
\begin{aligned}
MSE &= E[(x_{n+1} - \tilde{x}_{n+1})^2] \\
&= E(\varepsilon_{n+1}^2) \\
&= \sigma^2.
\end{aligned}
$$

(b) For the truncated version

$$
\begin{aligned}
\tilde{x}_{n+1}^n &= \beta_1 \tilde{\varepsilon}_n^n \\
&= \beta_1(\tilde{x}_n^n - \beta_1 \tilde{\varepsilon}_{n-1}^n) \\
&= \beta_1(x_n - \beta_1(\tilde{x}_{n-1}^n - \tilde{\varepsilon}_{n-2}^n)) \\
&= \ldots \\
&= \beta_1 x_n - \beta_1^2 x_{n-1} + \beta_1^3 x_{n-2} + \ldots + (-\beta_1)^n (\underbrace{\tilde{x}_1^n}_{=x_1} - \beta_1 \underbrace{\tilde{\varepsilon}_0^n}_{=0}) \\
&= -\sum_{j=1}^{n} (-\beta_1)^j x_{n-j+1}
\end{aligned}
$$

Computing the mean-square-error

$$
\begin{aligned}
MSE &= E[(x_{n+1} - \tilde{x}_{n+1}^{n})^2] \\
&= E[(\varepsilon_{n+1} - \sum_{j=n+1}^{\infty} (-\beta_1)^j x_{n-j+1})^2] \\
&= E[(\varepsilon_{n+1} - (-\beta_1)^{n+1} \underbrace{\sum_{j=n+1}^{\infty} (-\beta_1)^{j-(n+1)} x_{n-j+1})^2]}_{=\varepsilon_0}, \\
&= E[(\varepsilon_{n+1} - (-\beta_1)^{n+1}\varepsilon_0)^2] \\
&= E[\varepsilon_{n+1}^2] + \beta_1^{2(n+1)} E[\varepsilon_0^2] \\
&= \sigma^2 \left(1 + \beta_1^{2(n+1)}\right),
\end{aligned}
$$

Exercise 4 (Forecasting AR(1) processes)
The following AR(1) model is given:

$$
x_t = \alpha_1 x_{t-1} + \varepsilon_t
$$

similar to the previous exercise, determine the general form of the m-step ahead forecast x_{t+m}^t and show

$$
\mathbb{E}[(x_{t+m} - x_{t+m}^t)^2] = \sigma^2 \frac{1 - \alpha_1^{2m}}{1 - \alpha_1^2} \tag{14}
$$

The AR(1) process we can be written as (see problem set 2, exercise 2):

$$
x_{t+m} = \alpha_1^m x_t + \sum_{j=0}^{m-1} \alpha_1^j \varepsilon_{t+m-j}.
$$

Therefore, it follows for the forecast

$$
\begin{aligned}
\tilde{x}_{n+m} &= E(x_{n+m} | x_n, \dots, x_1) \\
&= \alpha_1^m \underbrace{E(x_n | x_n, \dots, x_1)}_{=x_n} + \sum_{j=0}^{m-1} \alpha_1^j \underbrace{E(\varepsilon_{n+m-j} | x_n, \dots, x_1)}_{=0} \\
&= \alpha_1^m x_n
\end{aligned}
$$

For the mean square error we have:

$$
\begin{aligned}
MSE = E\left[(x_{n+m} - \tilde{x}_{n+m})^2\right] &= E\left[(x_{n+m} - \alpha_1^m x_n)^2\right] \\
&= E\left[\left(\sum_{j=0}^{m-1} \alpha_1^j \varepsilon_{n+m-j}\right)^2\right]
\end{aligned}
$$

(because of the independence of the ε_t's and $E(\varepsilon_t) = 0$)

$$
\begin{aligned}
&= \sum_{j=0}^{m-1} \alpha_1^{2j} E(\varepsilon_{n+m-j}^2) \\
&= \sigma^2 \sum_{j=0}^{m-1} \alpha_1^{2j} \\
&= \sigma^2 \frac{1 - \alpha_1^{2m}}{1 - \alpha_1^2}
\end{aligned}
$$