means

$$\begin{aligned}
\epsilon_1 &= y_1 - \mu \\
\epsilon_2 &= y_2 - \mu - \theta\left(y_1 - \mu\right) \\
\epsilon_3 &= y_3 - \mu - \theta\left(y_2 - \mu - \theta\left(y_1 - \mu\right)\right) \\
\epsilon_4 &= y_4 - \mu - \theta\left(y_3 - \mu - \theta\left(y_2 - \mu - \theta\left(y_1 - \mu\right)\right)\right) \\
&\vdots \\
\epsilon_T &= y_T - \mu - \theta\left(\cdots\right)
\end{aligned}$$

- Substitute these back into the joint pdf. The joint pdf is now a function of the $y_t's$. (At this point, write in general functional form):

$$f\left(y_T, y_{T-1}, ..., y_1|\mu, \theta, \sigma_\epsilon^2\right)$$

Here, we stop interpreting the function $f()$ as the joint pdf because it is now a function of the **data.** Instead, we call it the **likelihood function.** PDFs are for random variables. Likelihood functions are for data. The $\mu, \theta, \sigma_\epsilon^2$ are parameters of the likelihood.

- Maximum likelihood estimation is done by asking the computer to search and those $\mu, \theta, \sigma_\epsilon^2$ that maximizes $f()$. We reduce the nonlinearity of the likelihood function by taking logs to get the **log likelihood**. Values that maximize the log-likelihood also maximize the likelihood.

- **Log likelihood**:

$$L\left(y_T, y_{T-1}, ..., y_1|\mu, \theta, \sigma_\epsilon^2\right) = \ln\left(f\left(y_T, y_{T-1}, ..., y_1|\mu, \theta, \sigma_\epsilon^2\right)\right)$$

- Let's look deeper. Let's take the logarithm of the joint pdf.

$$\ln\left(f\left(\cdot\right)\right) = -T\ln\left[\left(\sigma_\epsilon^2\right)^{\frac{1}{2}}\right] - T\ln\left[\left(2\pi\right)^{\frac{1}{2}}\right] - \frac{1}{2\sigma_\epsilon^2}\sum_{t=1}^{T}\epsilon_t^2$$

When we write the $\epsilon_t's$ in terms of the model, they become functions of the data y_t,and the logarithm of the joint pdf becomes the log likelihood function. Let's divide by T . Two things happen.

$$\frac{\ln\left(f\left(\cdot\right)\right)}{T} = -\ln\left[\left(\hat{\sigma}_\epsilon^2\right)^{\frac{1}{2}}\right] - \ln\left[\left(2\pi\right)^{\frac{1}{2}}\right] - \frac{1}{2\hat{\sigma}_\epsilon^2}\underbrace{\frac{1}{T}\sum_{t=1}^{T}\hat{\epsilon}_t^2}_{\hat{\sigma}_\epsilon^2}$$

First, σ_ϵ^2 becomes $\hat{\sigma}_\epsilon^2$, and second, ϵ_t becomes $\hat{\epsilon}_t$, (because ϵ_t can move around when we move the process parameters around), and third $\frac{1}{T}\sum_{t=1}^{T}\hat{\epsilon}_t^2 = \hat{\sigma}_\epsilon^2$. After cancellations, log likelihood becomes,

$$\begin{aligned}\frac{LL}{T} &= -\ln\left[\left(\hat{\sigma}_\epsilon^2\right)^{\frac{1}{2}}\right] - \ln\left[(2\pi)^{\frac{1}{2}}\right] - \frac{1}{2\hat{\sigma}_\epsilon^2}\hat{\sigma}_\epsilon^2 \\ &= \underbrace{-\frac{1}{2} - \ln\left[(2\pi)^{\frac{1}{2}}\right]}_{\text{constant}} - \ln\left[\left(\hat{\sigma}_\epsilon^2\right)^{\frac{1}{2}}\right]\end{aligned}$$

The constant terms don't matter when we are trying to maximize the function so it is common to write the log likelihood as

$$\frac{LL}{T} = -\ln\left[\left(\hat{\sigma}_\epsilon^2\right)^{\frac{1}{2}}\right] \tag{12}$$

So we choose θ, μ to minimize $\hat{\sigma}_\epsilon^2$.

6. Let's apply MA(1) to daily stock returns.

   `EviewsExamples/ARIMA_Models.wf1`

   ```
   Code:  equation eqma1.ls(optmethod=opg) djiaret c ma(1)
   ```

5.2.2 MA(2) model

1. Observations correlated with at most 2 lags of itself.

$$y_t = \mu_y + \epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2}$$

Next, you can verify that

$$\begin{aligned}E\left(y_t\right) &= \mu_y, \\ Var\left(y_t\right) &= \left(1+\theta_1^2+\theta_2^2\right)\sigma_\epsilon^2,\end{aligned}$$

$$\begin{aligned}Cov\left(y_t, y_{t-1}\right) &= \left(\theta_1+\theta_1\theta_2\right)\sigma_\epsilon^2, \\ \rho_1 &= \frac{\left(\theta_1+\theta_1\theta_2\right)\sigma_\epsilon^2}{\left(1+\theta_1^2+\theta_2^2\right)\sigma_\epsilon^2} = \frac{\left(\theta_1+\theta_1\theta_2\right)}{\left(1+\theta_1^2+\theta_2^2\right)} \\ Cov\left(y_t, y_{t-2}\right) &= \theta_2\sigma_\epsilon^2, \\ \rho_2 &= \frac{\theta_2\sigma_\epsilon^2}{\left(1+\theta_1^2+\theta_2^2\right)\sigma_\epsilon^2} = \frac{\theta_2}{\left(1+\theta_1^2+\theta_2^2\right)}\end{aligned}$$

and

$$Cov\left(y_t, y_{t-k}\right) = 0,$$

for $k > 2$.

2. **Forecasting** formula. Forecast deviation from mean.

$$\begin{aligned} E_t\left(\tilde{y}_{t+1}\right) &= E_t\left(\epsilon_{t+1} + \theta_1 \epsilon_t + \theta_2 \epsilon_{t-1}\right) \\ &= \theta_1 \epsilon_t + \theta_2 \epsilon_{t-1} \end{aligned}$$

$$E_t\left(\tilde{y}_{t+2}\right) = E_t\left(\epsilon_{t+2} + \theta_1 \epsilon_{t+1} + \theta_2 \epsilon_t\right) = \theta_2 \epsilon_t$$

$$E_t(\tilde{y}_{t+3}) = 0$$

Hence for any $k \geq 3 E_t(\tilde{y}_{t+k}) = 0$.

3. **Impulse response** function. Set all $\epsilon_{t-s} = 0$ for $s \neq 0$, $\epsilon_t = \sigma_\epsilon$, which we'll assume is $\sigma_\epsilon = 1$.

$$\begin{aligned} \tilde{y}_{t-1} &= 0 \\ \tilde{y}_t &= 1 \\ \tilde{y}_{t+1} &= \theta_1 \\ \tilde{y}_{t+2} &= \theta_2 \\ \tilde{y}_{t+3} &= 0 \end{aligned}$$

5.3 Autoregressive models

These are models of more durable or persistent dependence over time. Pure AR models can be estimated by least squares (actually, least squares and maximum likelihood are the same for AR models). Combined AR and MA models (ARMA) need to be estimated by maximum likelihood.