STAT 753: Stochastic Models and Simulations

Jaleesa Houle
University of Nevada, Reno - Spring 2024

Homework 2

Given data with the following monthly observations from Jan 1, 1974 to Jan 1, 2024:

DGS10: Market Yield on U.S. Treasury Securities at 10-Year Constant Maturity, Quoted on an Investment Basis, Percent, Monthly, Not Seasonally Adjusted.
MORTGAGE30US: 30-Year Fixed Rate Mortgage Average in the United States, Percent, Monthly, Not Seasonally Adjusted.
DGS30: Market Yield on U.S. Treasury Securities at 30-Year Constant Maturity, Quoted on an Investment Basis, Percent, Monthly, Not Seasonally Adjusted.
AAA: Moody's Seasoned Aaa Corporate Bond Yield, Percent, Monthly, Not Seasonally Adjusted.

1. Find the mean and the variance of each time series for the entire time period. Do the same for the common time period.

Solution / Code

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import statsmodels.api as sm
import statsmodels.formula.api as smf
from statsmodels.tsa.ar_model import AutoReg

df = pd.read_csv('rates.csv')
df.DGS30[df.DGS30=='.']=np.nan #change '.' entries to nan so they are easier to deal with in functions
df.DGS30 = df.DGS30.astype('float64')

print('Means for entire time period: \n \n', df[['DGS10','MORTGAGE30US', 'DGS30', 'AAA']].mean(), '\n \n', 'Variances for entire time period: \n \n', df[['DGS10','MORTGAGE30US', 'DGS30', 'AAA']].var())


>>Means for entire time period:
>>DGS10           5.974872
>>MORTGAGE30US    7.738512
>>DGS30           6.238880
>>AAA             7.135075
>>dtype: float64
>>
```

```
>>Variances for entire time period:
>>DGS10           10.568399
>>MORTGAGE30US    11.115838
>>DGS30            9.216175
>>AAA              8.367469
>>dtype: float64

print('Means for common time period: \n \n', df[['DGS10','MORTGAGE30US', 'DGS30', 'AAA']].dropna().mean(), '\n \n', 'Variances for common time period: \n \n', df[['DGS10','MORTGAGE30US', 'DGS30', 'AAA']].dropna().var())

>>Means for common time period:
>>DGS10           5.858648
>>MORTGAGE30US    7.652488
>>DGS30           6.238880
>>AAA             7.037780
>>dtype: float64
>>Variances for common time period:
>>DGS10           11.059629
>>MORTGAGE30US    11.740897
>>DGS30            9.216175
>>AAA              8.774280
>>dtype: float64
```

2. Regress 10-year interest rates upon itself with 1-month lag (AR of order 1). Show output. Analyze residuals for normality.

Solution / Code

```python
mod = AutoReg(df.DGS10, lags=1)
res = mod.fit()
print(res.summary())
```

```
                            AutoReg Model Results
==============================================================================
Dep. Variable:                  DGS10   No. Observations:                  601
Model:                     AutoReg(1)   Log Likelihood                -122.733
Method:               Conditional MLE   S.D. of innovations              0.297
Date:                Mon, 05 Feb 2024   AIC                            251.465
Time:                        22:08:00   BIC                            264.656
Sample:                             1   HQIC                           256.600
                                  601
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0189      0.025      0.743      0.457      -0.031       0.069
DGS10.L1       0.9960      0.004    267.071      0.000       0.989       1.003
                                    Roots
=============================================================================
                  Real          Imaginary           Modulus         Frequency
-----------------------------------------------------------------------------
AR.1            1.0040           +0.0000j            1.0040            0.0000
-----------------------------------------------------------------------------
```

Figure 1: Summary of 10-year interest rates regressed on itself with 1 month lag.

```python
fig = plt.figure(figsize=(16, 9))
fig = res.plot_diagnostics(lags=1, fig=fig)
print(scipy.stats.shapiro(res.resid))
>>ShapiroResult(statistic=0.946619987487793, pvalue=7.869244399985251e-14)
```

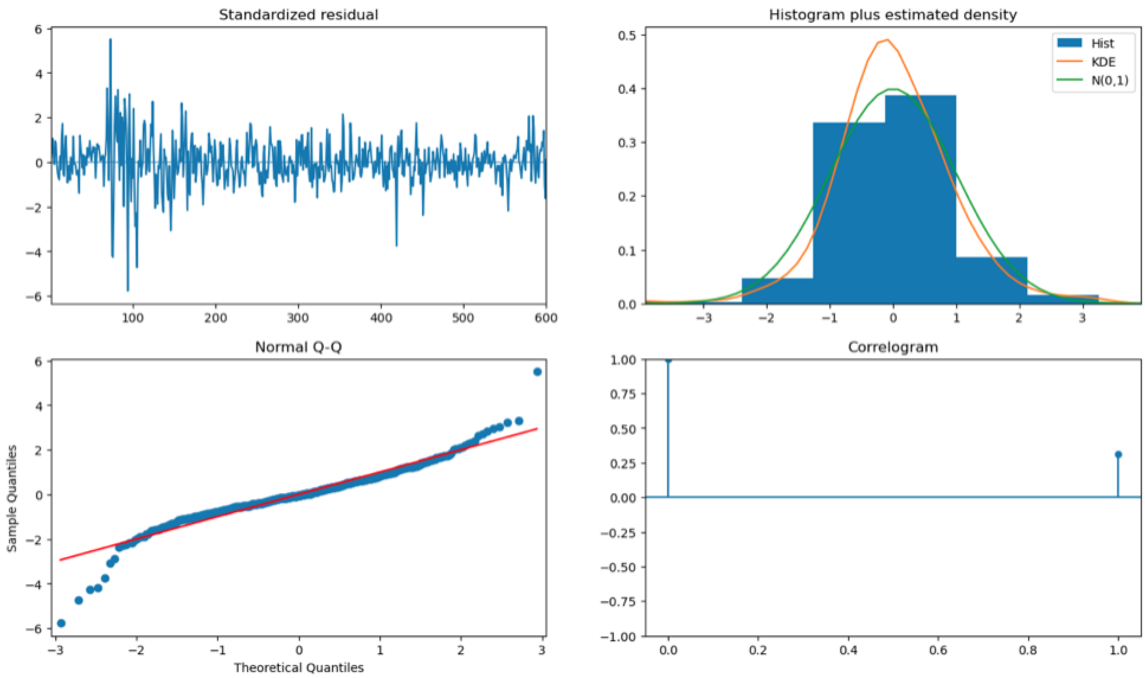

Figure 2: Residual diagnostic plots for AR model with 1 month lag.

From Figure 2, it is clear that there is some non-normality of residuals, as the variance of the standardized residuas is not constant, the histogram/density estimate does not quite match up to a normal distribution density curve, the QQ plot is a bit heavy tailed, and the correlogram shows that there is still a correlation of about 0.35 between data points on a 1 month lag. This non-normality is confirmed via a Shapiro-Wilk test of the residuals, which gives a p-value < 0.05, meaning that we reject the hypothesis that the residuals are from a normal distribution.

3. Regress 30-year mortgage rates upon three other rates. Show output. Analyze residuals for normality. Are all factors significant?

Solution / Code

```python
mod = smf.ols(formula='MORTGAGE30US ~ DGS10 + DGS30 + AAA', data=df.dropna())
res = mod.fit()
print(res.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:           MORTGAGE30US   R-squared:                       0.985
Model:                            OLS   Adj. R-squared:                  0.985
Method:                 Least Squares   F-statistic:                 1.207e+04
Date:                Mon, 05 Feb 2024   Prob (F-statistic):               0.00
Time:                        22:01:41   Log-Likelihood:                -313.26
No. Observations:                 563   AIC:                             634.5
Df Residuals:                     559   BIC:                             651.8
Df Model:                           3
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      1.0812      0.097     11.140      0.000       0.891       1.272
DGS10          1.1553      0.060     19.119      0.000       1.037       1.274
DGS30         -0.8816      0.077    -11.417      0.000      -1.033      -0.730
AAA            0.7535      0.057     13.205      0.000       0.641       0.866
==============================================================================
Omnibus:                      265.839   Durbin-Watson:                   0.303
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             1716.652
Skew:                           1.994   Prob(JB):                         0.00
Kurtosis:                      10.567   Cond. No.                         79.8
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Figure 3: Summary statistics of MLR.

```python
print(scipy.stats.shapiro(res.resid))

fig, ax =plt.subplots(nrows=1, ncols=2, figsize=(10,5), layout='tight')
plt.suptitle('MLS model: MORTGAGE30US ~ DGS10 + DGS30 + AAA')
slope, intercept = np.polyfit(res.fittedvalues, res.resid_pearson, 1)
abline_values = [slope * i + intercept for i in res.fittedvalues]
ax[0].scatter(res.fittedvalues, res.resid_pearson, s=10, rasterized=True)
ax[0].plot(res.fittedvalues, abline_values, 'r', rasterized=True )
ax[0].set_xlabel ('Fitted values')
ax[0].set_ylabel ('Standardized residuals')
sm.qqplot(res.resid_pearson, line='q', markersize=4, ax=ax[1], rasterized=True)
plt.show()

>>ShapiroResult(statistic=0.8606145977973938, pvalue=6.044019005410937e-22)
```

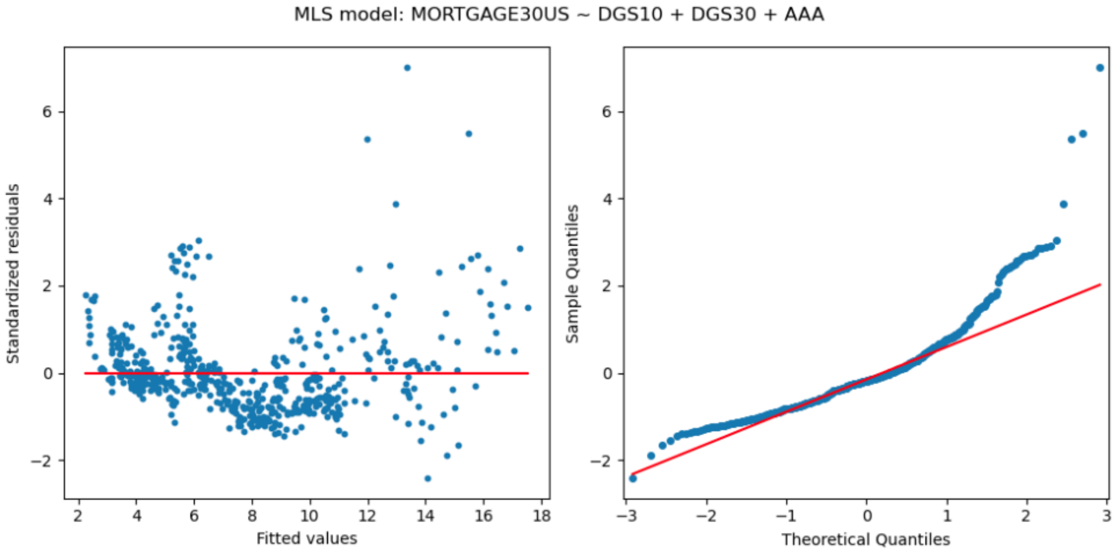

Figure 4: MLR regression residual diagnostic plots.

If we regress 30-year fixed mortgage rates on 10-year market yield on U.S. Treasury Securities, 30-year market yield on U.S. Treasury Securities, and Moody's Seasoned Aaa Corporate Bond Yield, we see that all three factors have a p-value < 0.05, so they are technically significant. However, there is clear heteroskedasticity of the residuals as demonstrated by Figure 4. A Shapiro-Wilk test confirms that the residuals are not from a normal distribution (p-value <0.05). This means that although all three variables are significant, the model is not a good fit to the data and therefore interpretation of the results may not be accurate, as underlying model assumptions are violated.

Theory

For any random variable X with finite fourth moment $E[X^4] < \infty$, define $k(X) = \frac{E[(X-E[X])^4]}{var(X)^2}$ the kurtosis.

1. Find the kurtosis of a normal random variable (with any mean and variance) using MGF.

Solution

By expanding the given function for $k(X)$, we get

$$k(X) = \frac{E[(X^4 - 4X^3E[X] + 6x^2E[X]^2 - 4XE[X]^3 + E[X]^4)]}{var(X)^2}. \tag{1}$$

The MGF of a normal random variable with mean μ and variance σ^2 is

$$M_X(t) = E[e^{(t\mu + \frac{1}{2}t^2\sigma^2)}] \tag{2}$$

where $E[X^i] = M_X^{(i)}(0)$ for $i = [1, 2, 3..]$.

To find the kurtosis using MGF, we will differentiate $M_X(t)$ four times in order to find expressions for $E[X], E[X^2], E[X^3]$ and $E[X^4]$. Doing so yields the following:

$$\begin{aligned} E[X] &= M_X^{'}(0) \\ &= \frac{d}{dt}\Big|_{t=0}(E[e^{(t\mu+\frac{1}{2}t^2\sigma^2)}]) \\ &= E[\frac{d}{dt}\Big|_{t=0}(e^{(t\mu+\frac{1}{2}t^2\sigma^2)})] \\ &= E[(\mu + \sigma^2 t)(e^{t\mu+\frac{1}{2}\sigma^2 t^2})]\Big|_{t=0} \\ &= \mu \end{aligned} \tag{3}$$

$$\begin{aligned} E[X^2] &= M_X^{''}(0) \\ &= E[\frac{d}{dt}\Big|_{t=0}(\mu + \sigma^2 t)(e^{t\mu+\frac{1}{2}\sigma^2 t^2})] \\ &= E[(\sigma^2 + (\mu + \sigma^2 t)^2)e^{t\mu+\frac{1}{2}\sigma^2 t^2}]\Big|_{t=0} \\ &= E[(\sigma^2 + \mu^2 + 2\mu\sigma^2 t + \sigma^4 t^2)e^{t\mu+\frac{1}{2}\sigma^2 t^2}]\Big|_{t=0} \\ &= E[(\sigma^2 + \mu^2] \\ &= \sigma^2 + \mu^2 \end{aligned} \tag{4}$$

$$\begin{aligned}
E[X^3] &= M_X'''(0) \\
&= E[\frac{d}{dt}\Big|_{t=0}((\sigma^2+\mu^2)+2\mu\sigma^2 t+\sigma^4 t^2)e^{t\mu+\frac{1}{2}\sigma^2 t^2}] \\
&= E[(\mu+\sigma^2 t)(\sigma^2+\mu^2)+(\mu+\sigma^2 t)(2\mu\sigma^2 t) \\
&+ 2\mu\sigma^2+(\sigma^4 t^2)(\mu+\sigma^2 t)+(2\sigma^4 t))e^{t\mu+\frac{1}{2}\sigma^2 t^2}]\Big|_{t=0} \\
&= E[(\mu)(\sigma^2+\mu^2)+2\mu\sigma^2] \\
&= \mu^3+3\sigma^2\mu
\end{aligned} \tag{5}$$

$$\begin{aligned}
E[X^4] &= M_X^4(0) \\
&= E[\frac{d}{dt}\Big|_{t=0}(\mu+\sigma^2 t)(\sigma^2+\mu^2)+(\mu+\sigma^2 t)(2\mu\sigma^2 t) \\
&+ 2\mu\sigma^2+(\sigma^4 t^2)(\mu+\sigma^2 t)+(2\sigma^4 t))e^{t\mu+\frac{1}{2}\sigma^2 t^2}]\Big|_{t=0} \\
&= E[\frac{d}{dt}\Big|_{t=0}(\sigma^6 t^3+2\sigma^4 t^2\mu+\mu^2\sigma^2 t \\
&+ 3\sigma^4 t+\mu\sigma^4 t^2+2\mu^2\sigma^2 t+\mu^3+3\mu\sigma^2)e^{t\mu+\frac{1}{2}\sigma^2 t^2}] \\
&= E[((\sigma^2 t+\mu)^2(\sigma^4 t^2+2\sigma^2\mu t)+(\sigma^6 t^2+2\sigma^4\mu t \\
&+ \mu^2\sigma^2+3\sigma^4)+(\sigma^2 t+\mu)(2\sigma^4 t+2\sigma^2\mu))e^{t\mu+\frac{1}{2}\sigma^2 t^2}\Big|_{t=0}] \\
&= E[(\sigma^8 t^4+4\sigma^6\mu t^3+(6\sigma^4\mu^2+6\sigma^6)t^2+(4\sigma^2\mu^3 \\
&+ 12\sigma^4\mu)t+(\mu^4+6\sigma^2\mu^2+3\sigma^4))e^{t\mu+\frac{1}{2}\sigma^2 t^2}\Big|_{t=0}] \\
&= \mu^4+6\sigma^2\mu^2+3\sigma^4
\end{aligned} \tag{6}$$

The variance can be computed as $var(X) = E[X^2] - E[X]^2 = \sigma^2 + \mu^2 - \mu^2 = \sigma^2$. By putting this all into the original equation, we get:

$$\begin{aligned}
k(X) &= \frac{E[(X^4-4X^3\mu+6X^2\mu^2-4X\mu^3+\mu^4)]}{(\sigma^2)^2} \\
&= \frac{E[X^4]-4E[X^3]\mu+6E[X^2]\mu^2-4E[X]\mu^3+\mu^4)}{\sigma^4} \\
&= \frac{(\mu^4+6\sigma^2\mu^2+3\sigma^4)-4(\mu^3+3\sigma^2\mu)\mu+6(\sigma^2+\mu^2)\mu^2-4\mu^4+\mu^4)]}{\sigma^4} \\
&= \frac{\mu^4+6\sigma^2\mu^2+3\sigma^4-4\mu^4-12\sigma^2\mu^2+6\sigma^2\mu^2+6\mu^4-4\mu^4+\mu^4)]}{\sigma^4} \\
&= \frac{3\sigma^4}{\sigma^4} \\
&= 3.
\end{aligned} \tag{7}$$

Thus, the kurtosis for a Gaussian random variable is 3.

2. Find the kurtosis of a Laplace random variable (with any mean and variance) using MGF.

Solution

The MGF of a Laplace random variable with mean μ and variance σ^2 is given as

$$M_X^i(t) = \frac{e^{\mu t}}{1 - \frac{\sigma}{\sqrt{2}}t^2}, \text{ for} |t| < \frac{\sqrt{2}}{\sigma}. \tag{8}$$

As above, we can use the MGF to determine $E[X]$, $E[X^2]$, $E[X^3]$ and $E[X^4]$. Doing so yields:

$$\begin{aligned} E[X] &= M_X^{'}(0) \\ &= E[\frac{d}{dt}\Big|_{t=0} \frac{e^{\mu t}}{1 - \frac{\sigma}{\sqrt{2}}t^2}] \\ &= E[\frac{\mu \mathrm{e}^{\mu t}}{1 - \frac{\sigma t^2}{\sqrt{2}}} + \frac{\sqrt{2}\,\sigma t \mathrm{e}^{\mu t}}{\left(1 - \frac{\sigma t^2}{\sqrt{2}}\right)^2}]\Big|_{t=0} \\ &= E[\mu] \\ &= \mu \end{aligned} \tag{9}$$

$$\begin{aligned} E[X^2] &= M_X^{''}(0) \\ &= E[\frac{d}{dt}\Big|_{t=0} \frac{\mu \mathrm{e}^{\mu t}}{1 - \frac{\sigma t^2}{\sqrt{2}}} + \frac{\sqrt{2}\,\sigma t \mathrm{e}^{\mu t}}{\left(1 - \frac{\sigma t^2}{\sqrt{2}}\right)^2}] \\ &= E[-\frac{\sqrt{2}\,(2\mu\sigma t - 2\sigma)\,\mathrm{e}^{\mu t}}{\left(\sigma t^2 - \sqrt{2}\right)^2} - \frac{\sqrt{2}\,\mu \cdot \left(\mu\sigma t^2 - 2\sigma t - \sqrt{2}\,\mu\right)\mathrm{e}^{\mu t}}{\left(\sigma t^2 - \sqrt{2}\right)^2} \\ &+ \frac{2^{\frac{5}{2}}\sigma t \cdot \left(\mu\sigma t^2 - 2\sigma t - \sqrt{2}\,\mu\right)\mathrm{e}^{\mu t}}{\left(\sigma t^2 - \sqrt{2}\right)^3}]\Big|_{t=0} \\ &= E[-\frac{\sqrt{2}\left(\mu^2\sigma^2 t^4 - 4\mu\sigma^2 t^3 + \left(6\sigma^2 - 2^{\frac{3}{2}}\mu^2\sigma\right)t^2 + 2^{\frac{5}{2}}\mu\sigma t + 2^{\frac{3}{2}}\sigma + 2\mu^2\right)\mathrm{e}^{\mu t}}{\left(\sigma t^2 - \sqrt{2}\right)^3}]\Big|_{t=0} \\ &= E[\frac{\sqrt{2}(2^{\frac{2}{3}}\sigma + 2\mu^2)}{\sqrt{2}^3}] \\ &= \sqrt{2}\sigma + \mu^2 \end{aligned} \tag{10}$$

$$
\begin{aligned}
E[X^3] &= M_X^{'''}(0) \\
&= E[\frac{d}{dt}\Big|_{t=0}(-\frac{\sqrt{2}\,(2\mu\sigma t-2\sigma)\,\mathrm{e}^{\mu t}}{\left(\sigma t^2-\sqrt{2}\right)^2}-\frac{\sqrt{2}\,\mu\cdot\left(\mu\sigma t^2-2\sigma t-\sqrt{2}\,\mu\right)\mathrm{e}^{\mu t}}{\left(\sigma t^2-\sqrt{2}\right)^2} \\
&+\frac{2^{\frac{5}{2}}\sigma t\cdot\left(\mu\sigma t^2-2\sigma t-\sqrt{2}\,\mu\right)\mathrm{e}^{\mu t}}{\left(\sigma t^2-\sqrt{2}\right)^3})] \\
&= E[-\frac{\sqrt{2}\,\mu\cdot\left(\mu^2\sigma^2t^4-4\mu\sigma^2t^3+\left(6\sigma^2-2^{\frac{3}{2}}\mu^2\sigma\right)t^2+2^{\frac{5}{2}}\mu\sigma t+2^{\frac{3}{2}}\sigma+2\mu^2\right)\mathrm{e}^{\mu t}}{\left(\sigma t^2-\sqrt{2}\right)^3} \\
&+\frac{3{\cdot}2^{\frac{3}{2}}\sigma t\cdot\left(\mu^2\sigma^2t^4-4\mu\sigma^2t^3+\left(6\sigma^2-2^{\frac{3}{2}}\mu^2\sigma\right)t^2+2^{\frac{5}{2}}\mu\sigma t+2^{\frac{3}{2}}\sigma+2\mu^2\right)\mathrm{e}^{\mu t}}{\left(\sigma t^2-\sqrt{2}\right)^4} \\
&-\frac{\sqrt{2}\left(4\mu^2\sigma^2t^3-12\mu\sigma^2t^2+2\left(6\sigma^2-2^{\frac{3}{2}}\mu^2\sigma\right)t+2^{\frac{5}{2}}\mu\sigma\right)\mathrm{e}^{\mu t}}{\left(\sigma t^2-\sqrt{2}\right)^3}]\Big|_{t=0} \\
&= 3\sqrt{2}\mu\sigma+\mu^3
\end{aligned} \tag{11}
$$

$$
\begin{aligned}
E[X^4] &= M_X^4(0) \\
&= E[\frac{d}{dt}\Big|_{t=0}-\frac{\sqrt{2}\,\mu\cdot\left(\mu^2\sigma^2t^4-4\mu\sigma^2t^3+\left(6\sigma^2-2^{\frac{3}{2}}\mu^2\sigma\right)t^2+2^{\frac{5}{2}}\mu\sigma t+2^{\frac{3}{2}}\sigma+2\mu^2\right)\mathrm{e}^{\mu t}}{\left(\sigma t^2-\sqrt{2}\right)^3} \\
&+\frac{3{\cdot}2^{\frac{3}{2}}\sigma t\cdot\left(\mu^2\sigma^2t^4-4\mu\sigma^2t^3+\left(6\sigma^2-2^{\frac{3}{2}}\mu^2\sigma\right)t^2+2^{\frac{5}{2}}\mu\sigma t+2^{\frac{3}{2}}\sigma+2\mu^2\right)\mathrm{e}^{\mu t}}{\left(\sigma t^2-\sqrt{2}\right)^4} \\
&-\frac{\sqrt{2}\left(4\mu^2\sigma^2t^3-12\mu\sigma^2t^2+2\left(6\sigma^2-2^{\frac{3}{2}}\mu^2\sigma\right)t+2^{\frac{5}{2}}\mu\sigma\right)\mathrm{e}^{\mu t}}{\left(\sigma t^2-\sqrt{2}\right)^3}] \\
&= E[-\frac{\sqrt{2}(\mu^4\sigma^4t^8-8\mu^3\sigma^4t^7+\left(36\mu^2\sigma^4-2^{\frac{5}{2}}\mu^4\sigma^3\right)t^6}{\left(\sigma t^2-\sqrt{2}\right)^5} \\
&+\frac{(3{\cdot}2^{\frac{7}{2}}\mu^3\sigma^3-96\mu\sigma^4)t^5+\left(120\sigma^4-15{\cdot}2^{\frac{5}{2}}\mu^2\sigma^3+12\mu^4\sigma^2\right)t^4-48\mu^3\sigma^2t^3}{\left(\sigma t^2-\sqrt{2}\right)^5} \\
&+\frac{(15{\cdot}2^{\frac{9}{2}}\sigma^3+24\mu^2\sigma^2-2^{\frac{7}{2}}\mu^4\sigma)t^2+\left(192\mu\sigma^2+2^{\frac{9}{2}}\mu^3\sigma\right)t+48\sigma^2+3{\cdot}2^{\frac{7}{2}}\mu^2\sigma+4\mu^4)\mathrm{e}^{\mu t}}{\left(\sigma t^2-\sqrt{2}\right)^5}]\Big|_{t=0} \\
&= 12\sigma^2+6\sqrt{2}\mu^2\sigma+\mu^4
\end{aligned} \tag{12}
$$

The variance can be computed as $var(X) = E[X^2]-E[X]^2 = \sqrt{2}\sigma+\mu^2-\mu^2 = \sqrt{2}\sigma$. Again, putting this all into the original equation, we get:

$$
\begin{aligned}
k(X) &= \frac{E[(X^4 - 4X^3\mu + 6X^2\mu^2 - 4X\mu^3 + \mu^4)]}{(\sqrt{2}\sigma)^2} \\
&= \frac{E[X^4] - 4E[X^3]\mu + 6E[X^2]\mu^2 - 4E[X]\mu^3 + \mu^4)}{2\sigma^2} \\
&= \frac{(12\sigma^2 + 6\sqrt{2}\mu^2\sigma + \mu^4) - 4(3\sqrt{2}\mu\sigma + \mu^3)\mu + 6(\sqrt{2}\sigma + \mu^2)\mu^2 - 4\mu^4 + \mu^4)]}{2\sigma^2} \\
&= \frac{(12\sigma^2 + 6\sqrt{2}\mu^2\sigma + \mu^4) - 4(3\sqrt{2}\mu\sigma + \mu^3)\mu + 6(\sqrt{2}\sigma + \mu^2)\mu^2 - 4\mu^4 + \mu^4)]}{2\sigma^2} \\
&= \frac{12\sigma^2 + 6\sqrt{2}\mu^2\sigma + \mu^4 - 12\sqrt{2}\mu^2\sigma - 12\sqrt{2}\mu^2\sigma - 4\mu^4 + 6\sqrt{2}\sigma\mu^2 + 6\mu^4 - 4\mu^4 + \mu^4}{2\sigma^2} \\
&= 6.
\end{aligned} \tag{13}
$$

Thus, the kurtosis for a Laplace random variable is 6.