时间序列分析
作业四

502022370071, 庄镇华, zhuangzh@lamda.nju.edu.cn

2022 年 12 月 9 日

作业提交注意事项

(1) 请严格参照教学立方网站所述提交作业，文件命名统一为学号 _ 姓名.pdf;

(2) 未按照要求提交作业，或提交作业格式不正确，将会被扣除部分作业分数；

(3) 除非有特殊情况（如因病缓交），否则截止时间后不接收作业，本次作业记零分。

1 Regression Model

1. An elasticity coefficient is the ratio of the percentage change in the forecast variable (y) to the percentage change in the predictor variable (x). Mathematically, the elasticity is defined as $(dy/dx) \times (x/y)$. Consider the log-log model,

$$\log y = \beta_0 + \beta_1 \log x + \varepsilon.$$

Express y as a function of x and show that the coefficient β_1 is the elasticity coefficient.

Solution. *(中英文皆可)*

$$\log y = \beta_0 + \beta_1 \log x + \varepsilon$$

根据链式法则，方程两边同时对 x 求导：

$$\frac{d \log y}{dy}\frac{dy}{dx} = \beta_1 \frac{d \log x}{dx}$$

$$\frac{1}{y}\frac{dy}{dx} = \beta_1 \frac{1}{x}$$

$$\frac{dy}{dx}\frac{x}{y} = \beta_1$$

即 $y = \exp(\beta_0 + \beta_1 \log x + \varepsilon)$ 且弹性系数为 β_1。

2. Using matrix notation it was shown that if $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where $\boldsymbol{\varepsilon}$ has mean $\mathbf{0}$ and variance matrix $\sigma^2\boldsymbol{I}$, the estimated coefficients are given by $\hat{\boldsymbol{\beta}} = \left(\boldsymbol{X}^\top\boldsymbol{X}\right)^{-1}\boldsymbol{X}^\top\boldsymbol{y}$ and a forecast is

given by $\hat{\boldsymbol{y}}=\boldsymbol{x}^{*}\hat{\boldsymbol{\beta}}=\boldsymbol{x}^{*}\left(\boldsymbol{X}^{\top}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\top}\boldsymbol{y}$ where $\boldsymbol{x}^{*}$ is a row vector containing the values of the predictors for the forecast (in the same format as $\boldsymbol{X}$), and the forecast variance is given by $\operatorname{Var}(\hat{y})=\sigma^{2}\left[1+\boldsymbol{x}^{*}\left(\boldsymbol{X}^{\top}\boldsymbol{X}\right)^{-1}\left(\boldsymbol{x}^{*}\right)^{\top}\right]$.

Consider the simple time trend model where $y_t=\beta_0+\beta_1 t+\epsilon_t$, ϵ_t has mean 0 and variance $\hat{\sigma}^2$. Using the following results,

$$\sum_{t=1}^{T} t=\frac{1}{2}T(T+1),\quad \sum_{t=1}^{T} t^{2}=\frac{1}{6}T(T+1)(2T+1)$$

derive the following expressions:

a. $\boldsymbol{X}^{\top}\boldsymbol{X}=\frac{1}{6}\begin{bmatrix}6T & 3T(T+1)\\ 3T(T+1) & T(T+1)(2T+1)\end{bmatrix}$

b. $\left(\boldsymbol{X}^{\top}\boldsymbol{X}\right)^{-1}=\frac{2}{T(T^{2}-1)}\begin{bmatrix}(T+1)(2T+1) & -3(T+1)\\ -3(T+1) & 6\end{bmatrix}$

c. $\hat{\beta}_0=\frac{2}{T(T-1)}\left[(2T+1)\sum_{t=1}^{T}y_t-3\sum_{t=1}^{T}ty_t\right]$
$\hat{\beta}_1=\frac{6}{T(T^{2}-1)}\left[2\sum_{t=1}^{T}ty_t-(T+1)\sum_{t=1}^{T}y_t\right]$

d. $\operatorname{Var}\left(\hat{y}_{T+h}\right)=\hat{\sigma}^{2}\left[1+\frac{2}{T(T-1)}\left(1-4T-6h+6\frac{(T+h)^{2}}{T+1}\right)\right]$

Solution. *(中英文皆可)*

a. 方程组 $y_t=\beta_0+\beta_1 t+\epsilon_t$（其中 $t=1,\cdots,T$）可以表示为

$$\begin{bmatrix}y_1\\ \vdots\\ y_t\\ \vdots\\ y_T\end{bmatrix}=\begin{bmatrix}1 & 1\\ \vdots & \vdots\\ 1 & t\\ \vdots & \vdots\\ 1 & T\end{bmatrix}\begin{bmatrix}\beta_0\\ \beta_1\end{bmatrix}+\begin{bmatrix}\varepsilon_1\\ \vdots\\ \varepsilon_t\\ \vdots\\ \varepsilon_T\end{bmatrix}$$

即 $\boldsymbol{X}=\begin{bmatrix}1 & 1\\ \vdots & \vdots\\ 1 & t\\ \vdots & \vdots\\ 1 & T\end{bmatrix}$, $\boldsymbol{X}^{T}\boldsymbol{X}=\begin{bmatrix}1 & \cdots & 1 & \cdots & 1\\ 1 & \cdots & t & \cdots & T\end{bmatrix}\begin{bmatrix}1 & 1\\ \vdots & \vdots\\ 1 & t\\ \vdots & \vdots\\ 1 & T\end{bmatrix}=\begin{bmatrix}T & \sum_{t=1}^{T}t\\ \sum_{t=1}^{T}t & \sum_{t=1}^{T}t^{2}\end{bmatrix}$

即 $\boldsymbol{X}^{\top}\boldsymbol{X}=\frac{1}{6}\begin{bmatrix}6T & 3T(T+1)\\ 3T(T+1) & T(T+1)(2T+1)\end{bmatrix}$

b. 设 $\left(\boldsymbol{X}^{\top}\boldsymbol{X}\right)^{-1}=\begin{bmatrix}a & b\\ c & d\end{bmatrix}$，则 $\begin{cases}6Ta+3T(T+1)c=1\\ 6Tb+3T(T+1)d=0\\ 3T(T+1)a+T(T+1)(2T+1)c=0\\ 3T(T+1)b+T(T+1)(2T+1)d=1\end{cases}$

解得 $\begin{cases} a = \frac{2(2T+1)}{T(T-1)} \\ b = -\frac{6}{T(T-1)} \\ c = -\frac{6}{T(T-1)} \\ d = \frac{12}{T(T^2-1)} \end{cases}$ ，即 $\left(\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1} = \frac{2}{T(T^2-1)} \begin{bmatrix} (T+1)(2T+1) & -3(T+1) \\ -3(T+1) & 6 \end{bmatrix}$

c.

$$
\begin{aligned}
\hat{\boldsymbol{\beta}} &= \left(\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1} \boldsymbol{X}^\top \boldsymbol{y} \\
&= \frac{2}{T\left(T^2-1\right)} \begin{bmatrix} (T+1)(2T+1) & -3(T+1) \\ -3(T+1) & 6 \end{bmatrix} \begin{bmatrix} 1 & \cdots & 1 & \cdots & 1 \\ 1 & \cdots & t & \cdots & T \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_t \\ \vdots \\ y_T \end{bmatrix} \\
&= \frac{2}{T\left(T^2-1\right)} \begin{bmatrix} (T+1)(2T+1) & -3(T+1) \\ -3(T+1) & 6 \end{bmatrix} \begin{bmatrix} \sum_{t=1}^T y_t \\ \sum_{t=1}^T t y_t \end{bmatrix} \\
&= \frac{2}{T\left(T^2-1\right)} \begin{bmatrix} (T+1)(2T+1) \sum_{t=1}^T y_t - 3(T+1) \sum_{t=1}^T t y_t \\ -3(T+1) \sum_{t=1}^T y_t + 6 \sum_{t=1}^T t y_t \end{bmatrix} \\
&= \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix}
\end{aligned}
$$

即 $\hat{\beta}_0 = \frac{2}{T(T-1)} \left[(2T+1) \sum_{t=1}^T y_t - 3 \sum_{t=1}^T t y_t\right]$,

$\hat{\beta}_1 = \frac{6}{T(T^2-1)} \left[2 \sum_{t=1}^T t y_t - (T+1) \sum_{t=1}^T y_t\right]$。

d. $y_{T+h} = \hat{\beta}_0 + \hat{\beta}_1(T+h) + \epsilon_{T+h}$, ϵ_{T+h} 均值为 0, 方差为 $\hat{\sigma}^2$, 即 $\boldsymbol{x}_{T+h}^* = \begin{bmatrix} 1 & T+h \end{bmatrix}$,

$$
\begin{aligned}
\operatorname{Var}\left(\hat{y}_{T+h}\right) &= \hat{\sigma}^2 \left[1 + \boldsymbol{x}_{T+h}^* \left(\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1} \left(\boldsymbol{x}_{T+h}^*\right)^\top\right] \\
&= \hat{\sigma}^2 \left[1 + \frac{2}{T\left(T^2-1\right)} \begin{bmatrix} 1 & T+h \end{bmatrix} \begin{bmatrix} (T+1)(2T+1) & -3(T+1) \\ -3(T+1) & 6 \end{bmatrix} \begin{bmatrix} 1 \\ T+h \end{bmatrix}\right] \\
&= \hat{\sigma}^2 \left[1 + \frac{2}{T\left(T^2-1\right)} \begin{bmatrix} 1 & T+h \end{bmatrix} \begin{bmatrix} (T+1)(2T+1) - 3(T+1)(T+h) \\ 6(T+h) - 3(T+1) \end{bmatrix}\right] \\
&= \hat{\sigma}^2 \left[1 + \frac{2}{T\left(T-1\right)} \left[(2T+1) - 3(T+h) + 6\frac{(T+h)^2}{T+1} - 3(T+h)\right]\right] \\
&= \hat{\sigma}^2 \left[1 + \frac{2}{T(T-1)} \left(1 - 4T - 6h + 6\frac{(T+h)^2}{T+1}\right)\right]
\end{aligned}
$$

2 ARIMA Models

1. In this exercise, we experiment with the python implementation of ARIMA in Darts. Consider the AUS Airpassengers dataset (see aus_airpassengers.csv), the total number of passengers (in millions) from Australian air carriers for the period 1970-2011.

   a. Use AutoARIMA to find an appropriate ARIMA model. What model was selected? Check that the residuals look like white noise. Plot forecasts for the next 10 periods.

b. Write the model in terms of the backshift operator.

c. Plot forecasts from an ARIMA $(0, 1, 0)$ model with a linear trend and compare these to part a.

d. Plot forecasts from an ARIMA $(2, 1, 2)$ model with a linear trend and compare these to parts a and c. Remove the trend and see what happens.

e. Plot forecasts from an ARIMA $(0, 2, 1)$ model with a constant. What happens?

Solution. *(中英文皆可)*

a. 自动选择的模型是 *ARIMA(0,2,1)*，模型残差如图*2*所示，可以认为其近似于白噪声，使用 *statsmodels.stats.diagnostic* 中的 *Ljung-Box* 检验是否是白噪声，得到的 p 值均大于 *0.05*，因此接受原假设。接下来十年的预测图像如图*3*所示。

图 1: AutoARIMA 模型参数图

SARIMAX Results

Dep. Variable:	y	No. Observations:	47
Model:	SARIMAX(0, 2, 1)	Log Likelihood	-97.019
Date:	Fri, 09 Dec 2022	AIC	198.038
Time:	19:23:16	BIC	201.651
Sample:	0	HQIC	199.385
	- 47		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ma.L1	-0.8963	0.114	-7.842	0.000	-1.120	-0.672
sigma2	4.2120	0.420	10.023	0.000	3.388	5.036

Ljung-Box (L1) (Q):	1.48	Jarque-Bera (JB):	104.43
Prob(Q):	0.22	Prob(JB):	0.00
Heteroskedasticity (H):	20.02	Skew:	1.61
Prob(H) (two-sided):	0.00	Kurtosis:	9.73

图 2: AutoARIMA 模型残差图

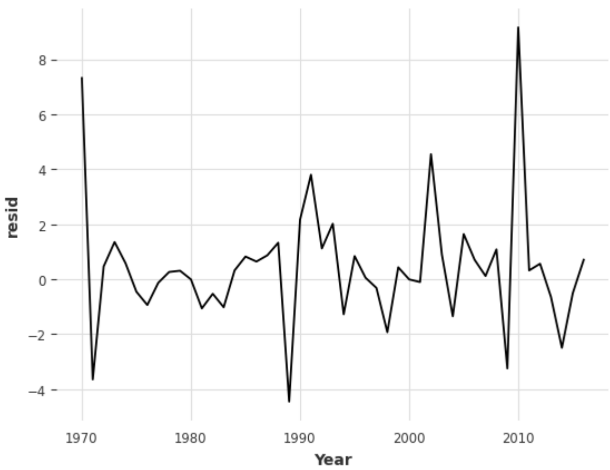
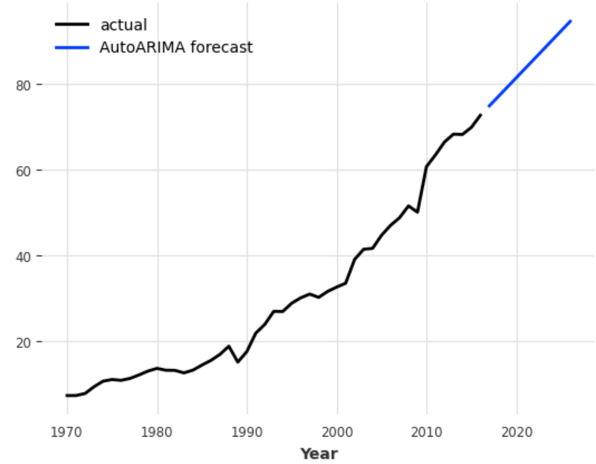


图 3: AutoARIMA 模型预测图



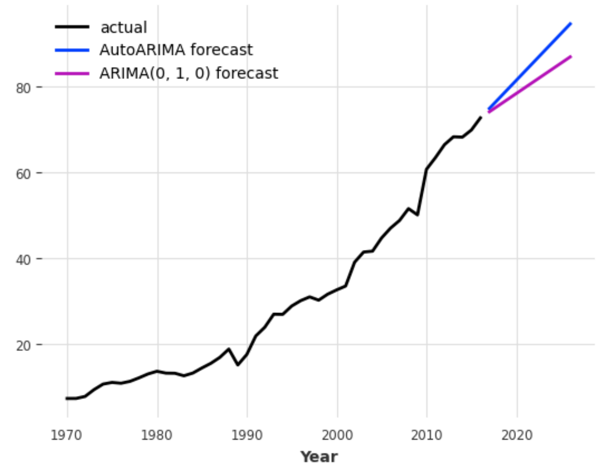
b. 参考模型参数图*1*，其中 *ma.L1* 表示 ε_{t-1} 的系数，*sigma2* 表示 ε_t 的方差。因此，*AutoARIMA* 选择的模型可以被表示为

$$(1-B)^2 Y_t = (1-0.8963B)\varepsilon_t$$

其中 $\varepsilon_t \sim \mathcal{N}(0, 4.2120)$。

c. 带有线性趋势的 *ARIMA(0,1,0)* 模型和 *AutoARIMA* 模型的预测对比图如图*4*所示。可以看到带有线性趋势的 *ARIMA(0,1,0)* 模型的预测值要低于 *AutoARIMA* 模型的预测值，这可能是因为 *AutoARIMA* 模型需要经过两次差分，而 *ARIMA(0,1,0)* 仅仅需要一次差分，*AutoARIMA* 模型还原时累积的差值要更大一些。

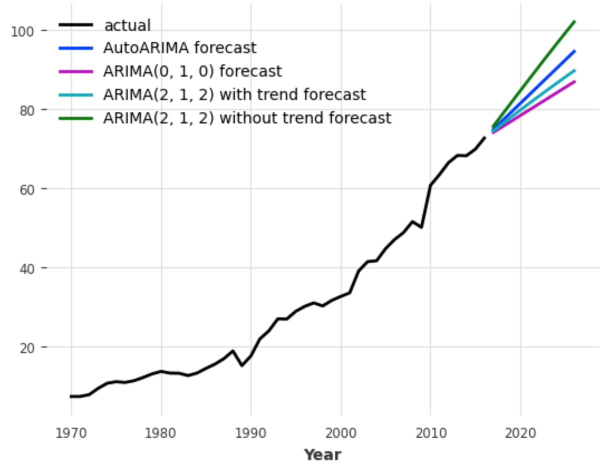
图 4: ARIMA(0,1,0) 模型预测对比图



d. 带有线性趋势项、没有线性趋势项的 *ARIMA(2,1,2)* 模型和之前模型的预测对比图如图*5*所示。可以看到去掉线性趋势项之后，模型预测的值偏高，在所有模型中是预测值最高的，这可能是因为带有线性趋势项的 *ARIMA(2,1,2)* 模型相当于对一次差分后的

结果加入常数，去掉线性趋势项后，一次差分的差值比之前要大一些，累加的时候最后结果会比之前更大。

图 5: ARIMA(2,1,2) 模型预测对比图



e. 无论是加入常数趋势项还是线性趋势项，代码都报错。因为这样做是无意义的，即当模型的 $d = 2$ 时，无论是加入常数趋势项还是线性趋势项，在两次差分后，常数趋势项和线性趋势项都会被消掉，因此是无意义的，代码会报错，无法生成图像。

Appendix

```python
import pandas as pd
import darts
from darts.models import forecasting
# 读取数据
df = pd.read_csv("aus_airpassengers.csv", index_col=0)
series = darts.TimeSeries.from_dataframe(df, "Year", "Passengers")

# 不同模型预测比较
arima = forecasting.auto_arima.AutoARIMA()
model = arima.fit(series = series)
forecast = model.predict(10)
series.plot(label="actual")
forecast.plot(label="AutoARIMA forecast")

arimac = forecasting.arima.ARIMA(p=0, d=1, q=0, trend='t')
modelc = arimac.fit(series = series)
forecastc = modelc.predict(10)
forecastc.plot(label="ARIMA(0, 1, 0) forecast")

arimad = forecasting.arima.ARIMA(p=2, d=1, q=2, trend='t')
modeld = arimad.fit(series = series)
forecastd = modeld.predict(10)
forecastd.plot(label="ARIMA(2, 1, 2) with trend forecast")

arimadd = forecasting.arima.ARIMA(p=2, d=1, q=2, trend=None)
modeldd = arimadd.fit(series = series)
forecastdd = modeldd.predict(10)
forecastdd.plot(label="ARIMA(2, 1, 2) without trend forecast")

# 查看模型参数
modelc.model.summary()

# 查看残差
from matplotlib import pyplot as plt
res = model.model.model_.resid()
plt.xlabel('Year')
plt.ylabel('resid')
plt.plot(df['Year'], res)

# 判断是否属于白噪声
from statsmodels.stats.diagnostic import acorr_ljungbox
ljungbox_result = acorr_ljungbox(res, lags=20)  # 返回统计量和p值，lags为检验的延迟数
```