INDR 450/550

Spring 2022

Lecture 11: Regression for Time Series

March 21, 2022

Fikri Karaesmen

Announcements

- Class Exercise at the end of lecture today. If you are participating online, please upload your document under Course Contents/Class Exercises
- HW 2 available soon.
- The first four labs were uploaded. Please follow them.
    - Next HW based on lab2 and lab3

Q&A

- Consider the AR(1) model: $Y_t = \mathrm{c} + \phi_1 \, Y_{t-1} + \epsilon_t$
- And the linear regression model:

$$y_t = \beta_0 + \beta_1 \, y_{t-1} + \epsilon_t$$

- They look very similar but
  - Note that the AR(1) model is a stochastic process, we can compute the probability distribution of Y_t conditional on past observations.
  - The regression typically is not concerned with the evolution of the predictors. We don't have a complete model of how the system evolves and where it will evolve to in the future.

Q&A

- The AR(1) model: $Y_t = \mathrm{c} + \phi_1 \, Y_{t-1} + \epsilon_t$
- And the linear regression model:

$$y_t = \beta_0 + \beta_1 \, y_{t-1} + \epsilon_t$$

- However, in terms of fitting the parameters from data, they are similar:
  - AR model fits parameters using MLE.
  - OLS regression fits the parameters based on squared residual minimization.
  - The resulting fits should be similar.

Q&A

- But there is no such easy correspondence for MA terms. This is why the general estimation tool for ARIMA processes is MLE.
- ARIMA processes perform quite well despite the restrictive assumptions.
- Regression allows more flexibility in modeling.

Summary of ACF and PACF patterns for simple AR and MA models

Process	ACF	PACF
AR(1)	Exponential decay: on positive side if $\phi_1 > 0$ and alternating in sign starting on negative side if $\phi_1 < 0$.	Spike at lag 1, then cuts off to zero: spike positive if $\phi_1 > 0$, negative if $\phi_1 < 0$.
AR(p)	Exponential decay or damped sine-wave. The exact pattern depends on the signs and sizes of $\phi_1, \ldots, \phi_p$.	Spikes at lags 1 to p, then cuts off to zero.
MA(1)	Spike at lag 1 then cuts off to zero: spike positive if $\theta_1 < 0$, negative if $\theta_1 > 0$.	Exponential decay: on negative side if $\theta_1 > 0$ and alternating in sign starting on positive side if $\theta_1 < 0$.
MA(q)	Spikes at lags 1 to q, then cuts off to zero.	Exponential decay or damped sine-wave. The exact pattern depends on the signs and sizes of $\theta_1, \ldots, \theta_q$.

Table 7-2: *Expected patterns in the ACF and PACF for simple AR and MA models.*

Please note that the MA-terms are defined with a negative sign in this Reference. This is why the signs are reversed in the examples in Lab 3.

$$Y_t = c - \theta_1 \varepsilon_{t-1} - \theta_2\, \varepsilon_{t-2} + \varepsilon_t$$

Makridakis, Wheelwright and Hyndman (1997)

Regression for Time Series

- Consider the following linear model:

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \ldots + \beta_n x_{nt} + \epsilon_t$$

- y_t is the forecast and x_{kt} are the predictors.
- We are therefore looking for a linear relationship between the predictors and the response (the forecast).
- Note that in the setting of forecasting, this is somewhat different than designing a controlled experiment where we can control the levels of the predictors. The predictors that are available to us cannot be controlled in general.

Regression for Time Series: Basic Predictors: Categorical Variables

- We can use dummies to mark months of the year, quarters of the year, hours of the day etc.
- We can also use dummies to mark irregular (non-seasonal) exceptions (holidays, days of Ramadan, promotions, school holidays etc.)
- This is great but note that we may easily end up with a very large number of dummies!

Regression for Time Series: Basic Predictors: Australian Beer Production

- The Australian Beer Production Data is strongly seasonal. We can try to fit:

$$y_t = \beta_0 + \beta_1 t + \beta_2 x_{1t} + \beta_3 x_{2t} + \ldots + \beta_{12} x_{11,t} + \epsilon_t$$

- where $x_{1t}, \ldots, x_{11,t}$ are the monthly dummies.

In [16]:
```
df = pd.read_csv('ausbeer_dummies.csv', index_col=0)
df.head()
```

Out[16]:

Month	Production	t	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11
1	164	1	1	0	0	0	0	0	0	0	0	0	0
2	148	2	0	1	0	0	0	0	0	0	0	0	0
3	152	3	0	0	1	0	0	0	0	0	0	0	0
4	144	4	0	0	0	1	0	0	0	0	0	0	0
5	155	5	0	0	0	0	1	0	0	0	0	0	0

Regression for Time Series: Basic Predictors: Australian Beer Production

$$y_t = \beta_0 + \beta_1 t + \beta_2 x_{1t} + \beta_3 x_{2t} + \ldots + \beta_{12} x_{11,t} + \epsilon_t$$

Dep. Variable:	Production	**R-squared:**	0.836
Model:	OLS	**Adj. R-squared:**	0.791
Method:	Least Squares	**F-statistic:**	18.30
Date:	Tue, 08 Mar 2022	**Prob (F-statistic):**	3.97e-13
Time:	12:20:51	**Log-Likelihood:**	-194.93
No. Observations:	56	**AIC:**	415.9
Df Residuals:	43	**BIC:**	442.2
Df Model:	12		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	192.9750	5.015	38.477	0.000	182.861	203.089
t	-0.2158	0.075	-2.887	0.006	-0.367	-0.065
M1	-39.7792	6.030	-6.597	0.000	-51.940	-27.619
M2	-48.5633	6.026	-8.059	0.000	-60.715	-36.411
M3	-30.9475	6.023	-5.139	0.000	-43.093	-18.802
M4	-46.7317	6.020	-7.762	0.000	-58.873	-34.591
M5	-46.1158	6.019	-7.662	0.000	-58.254	-33.978
M6	-58.5000	6.018	-9.720	0.000	-70.637	-46.363
M7	-51.8842	6.019	-8.620	0.000	-64.022	-39.746
M8	-42.2683	6.020	-7.021	0.000	-54.409	-30.127
M9	-46.6475	6.348	-7.349	0.000	-59.449	-33.846
M10	-19.6817	6.346	-3.102	0.003	-32.479	-6.884
M11	-2.9658	6.344	-0.467	0.643	-15.760	9.829

Regression for Time Series: Basic Predictors: Australian Beer Production

$$y_t = \beta_0 + \beta_1 t + \beta_2 x_{1t} + \beta_3 x_{2t} + \ldots + \beta_{12} x_{11,t} + \epsilon_t$$

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	192.9750	5.015	38.477	0.000	182.861	203.089
t	-0.2158	0.075	-2.887	0.006	-0.367	-0.065
M1	-39.7792	6.030	-6.597	0.000	-51.940	-27.619
M2	-48.5633	6.026	-8.059	0.000	-60.715	-36.411
M3	-30.9475	6.023	-5.139	0.000	-43.093	-18.802
M4	-46.7317	6.020	-7.762	0.000	-58.873	-34.591
M5	-46.1158	6.019	-7.662	0.000	-58.254	-33.978
M6	-58.5000	6.018	-9.720	0.000	-70.637	-46.363
M7	-51.8842	6.019	-8.620	0.000	-64.022	-39.746
M8	-42.2683	6.020	-7.021	0.000	-54.409	-30.127
M9	-46.6475	6.348	-7.349	0.000	-59.449	-33.846
M10	-19.6817	6.346	-3.102	0.003	-32.479	-6.884
M11	-2.9658	6.344	-0.467	0.643	-15.760	9.829

Our prediction for month 4 is: 192.98 -0.2158 (4) -46.73

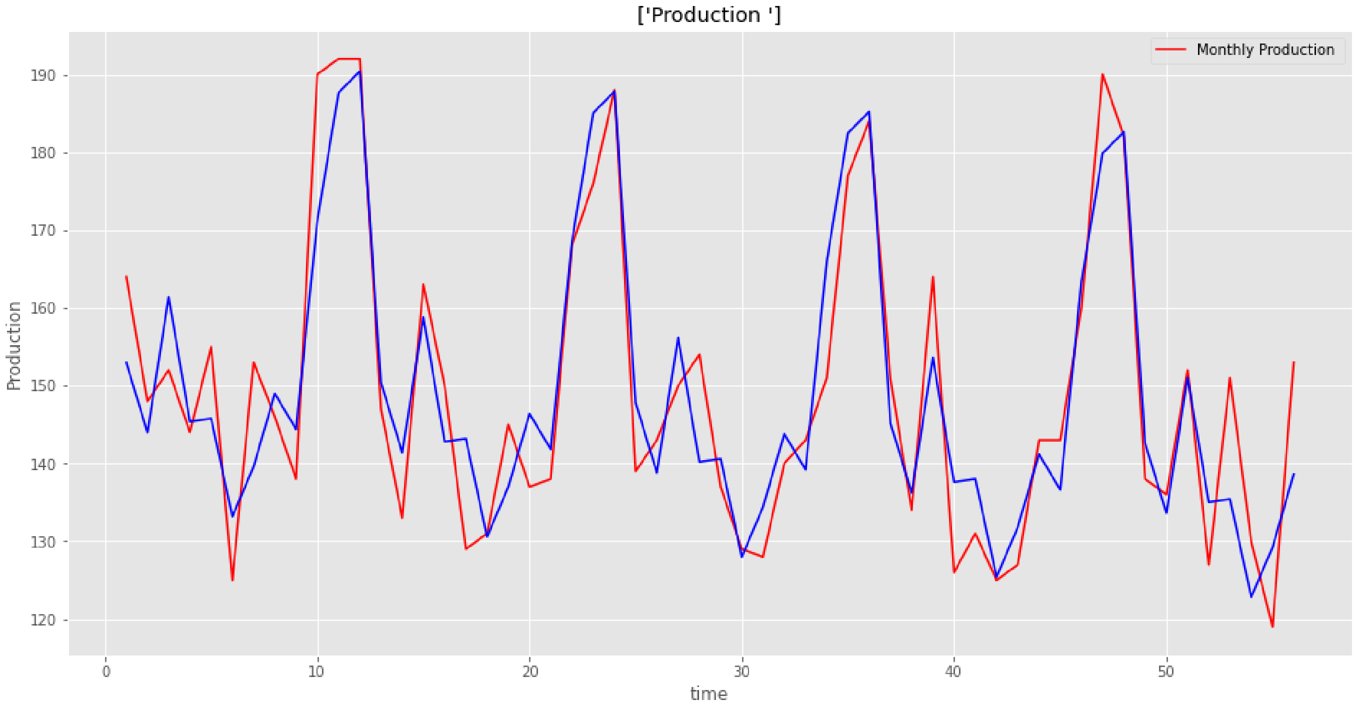
Our prediction for month 11 is: 192.98 -0.2158 (11) – 2.97

Our prediction for month 12 is: 192.98 -0.2158 (12)

Our prediction for month 26 is: 192.98 -0.2158 (26) -48.56

Month 12 is clearly the peak month for sales, all other months have negative seasonality factors wrt to month 12.

Regression for Time Series: Basic Predictors: Australian Beer Production



In-sample predictions in blue, and the observed production in red.

```
In [9]: error_beer = prod - result_beer.fittedvalues
```

```
In [10]: mse_beer = np.mean(np.square(error_beer ))
rmse_beer = np.sqrt(mse_beer)
mae_beer = np.mean(np.abs(error_beer ))
mape_beer = np.mean(np.abs(error_beer )/prod)
print('MSE Beer = ', mse_beer)
print('RMSE Beer = ', rmse_beer)
print('MAE Beer = ', mae_beer)
print('MAPE Beer = ', mape_beer)
```

```
MSE Beer =  61.805178571428556
RMSE Beer =  7.861626967201417
MAE Beer =  6.441249999999998
MAPE Beer =  0.04378739163608506
```

Regression for Time Series: Some Issues

- Some Issues
- 1. Non-linearities
- The relationship between the predictor x and the response y may not be linear.
- If we can figure out the relationship, we saw that we can transform the response as needed (i.e. take $\sqrt{x}$, x^3 or $\log x$ instead of x).
- There are other useful non-linear transformations (we'll see those later if time permits)

Regression for Time Series: Some Issues

- 2. Correlation of Error Terms
- We hope that ϵ_t does not provide any information on ϵ_{t+1}.
- It's a good idea to check the ACF and the PACF of the residuals.
- If auto-correlation shows, we should consider fitting an ARIMA model to the residuals.

Regression for Time Series: Some Issues

- 3. Non-constant Variance of Error Terms
- Non-constant variance is known as heteroscedasticity
- This would show as increasing fluctuations on the residual plot.
- We have seen that taking a log or square root transformation of the response y helps.

Regression for Time Series: Some Issues

- 4. Outliers
- These are points where the prediction $\hat{y}_t$ is far from the observation y_t.
- There are many reasons: inaccurate recording of data, major exceptions (like the start of a pandemic or a war, rapid price changes).
- We can distinguish major outliers from the residual plot.
- We can remove those outliers that are due to recording errors or exceptional situations by removing them or averaging them by similar responses.
- However, we don't want to be removing outliers due to predictors that can be foreseen (i.e. start of ramadan).

Regression for Time Series: Some Issues

- 5. High Leverage Points
- These are points where the predictor $\hat{x}_{it}$ takes an exceptional value with respect to its average.
- High leverage points have a big impact on the regression line.
- This is a tricky situation especially with multiple predictors.
- There is a leverage statistic (see Chapter 3 of James et al.)

Regression for Time Series: Some Issues

- 6. Collinearity
- Two or more predictors being strongly related.
- This is a big issue because we would not able to completely distinguish the individual effects such predictors in a linear regression.
- Checking for pairwise collinearity (i.e. between two predictors) is easy by computing the correlation matrix of the responses. We can then remove the variables that are strongly correlated.
- Multi-collinearity is trickier
- Collinearity reduces the accuracy of the estimates and causes the standard error to grow. This reduces the power of the standard hypothesis test: $H_0 = \beta_k = 0$.

Regression for Time Series: K – Nearest Neighbours

- Comparison to a non-parametric method: k-nearest neigbours (KNN) regression
- Let's say we would like to make a prediction for the point x_{1t} using the data available up to time $t-1$.
- We identify the K-nearest points to x_{1t} among $\{x_{11}, x_{12}, ...x_{1,t-1}\}$. Let us say that the nearest points are those in the set $\mathcal{N}_l$. Then our prediction is simply:

$$\hat{y}_t = \frac{\sum_{x_{i\tau} \in \mathcal{N}_l} y_{\tau+1}}{K}$$

- If we take $K = 1$, our prediction is given by the response to the nearest predictor in the training set.
- If we take $K = 10$ we are smoothing the prediction over the 10 nearest predictors.

Regression for Time Series: K – Nearest Neighbours

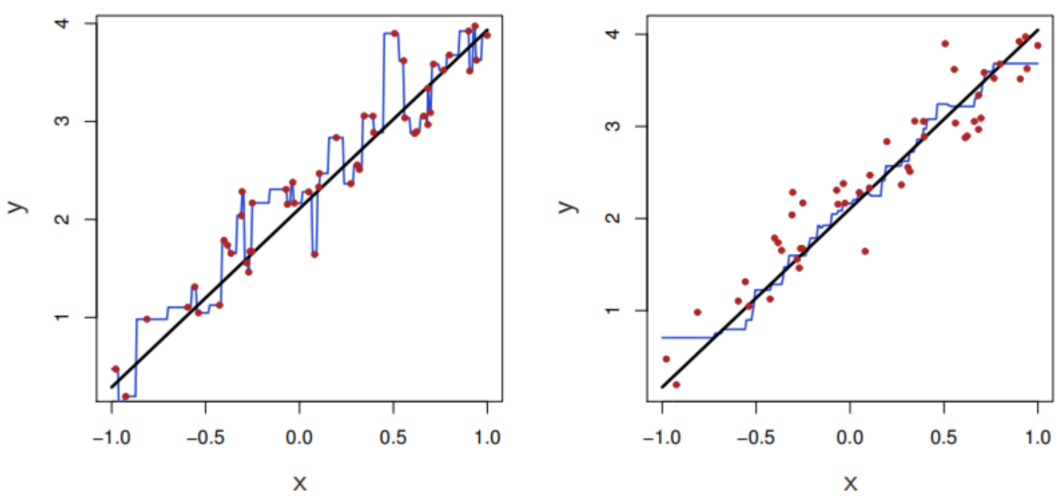


FIGURE 3.17. *Plots of $\hat{f}(X)$ using KNN regression on a one-dimensional data set with 50 observations. The true relationship is given by the black solid line.* Left: *The blue curve corresponds to $K = 1$ and interpolates (i.e. passes directly through) the training data.* Right: *The blue curve corresponds to $K = 9$, and represents a smoother fit.*

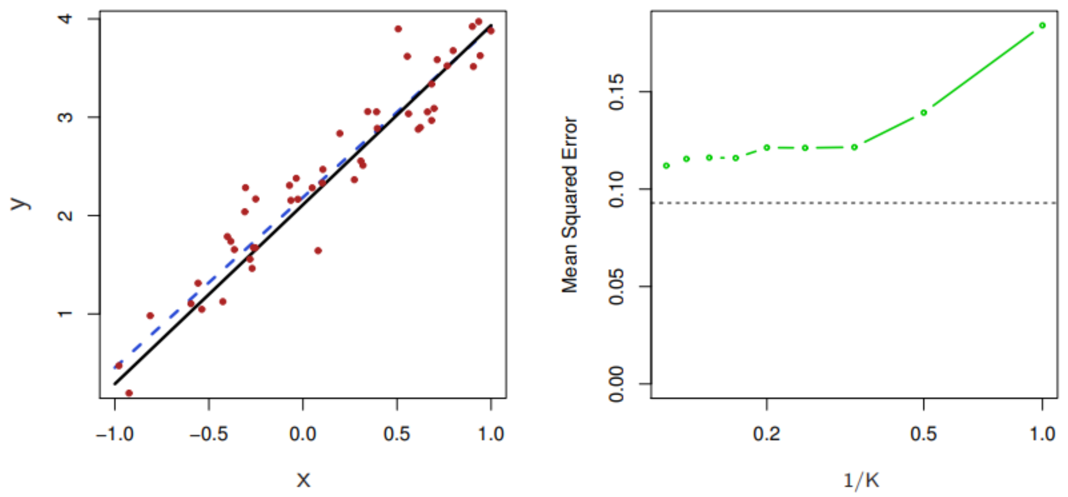


FIGURE 3.18. *The same data set shown in Figure 3.17 is investigated further.* Left: *The blue dashed line is the least squares fit to the data. Since $f(X)$ is in fact linear (displayed as the black line), the least squares regression line provides a very good estimate of $f(X)$.* Right: *The dashed horizontal line represents the least squares test set MSE, while the green solid line corresponds to the MSE for KNN as a function of $1/K$ (on the log scale). Linear regression achieves a lower test MSE than does KNN regression, since $f(X)$ is in fact linear. For KNN regression, the best results occur with a very large value of K, corresponding to a small value of $1/K$.*

Figures from James et al.

Regression for Time Series: K – Nearest Neighbours: Example

- Here are some basic issues for time-series regression.
- If there is a strong trend in the data, the nearest neighbours are likely to be the most recent observations.
  - A KNN regression would then be very similar to a simple Moving Avarege forecast with K periods.
  - The forecasts will always lag behind. Clearly, data will have to be detrended before implementing a KNN- regression.
- Seasonality with trend may not be so bad.
- Simple auto-correlation structure can be captured.

Regression for Time Series: K – Nearest Neighbours: Example

- Here's a seasonal example. Consider the Australian Beer Production Data. Let us say that we would like to predict the demand in month 48 based the first 46 months.
- In month 47, the production was 190 units.
- To implement 1-NN, we search in months 1 to 46, the month where the production was nearest to 190
  - In month, the production was also 190 units (month 10 is the nearest neigbour). In the following month (month 11), production was 192 units.
  - Therefore, our 1-NN estimate for the demand in period 48 is 192.

41	131
42	125
43	127
44	143
45	143
46	160
47	190
48	182
49	138
50	136
51	152
52	127
53	151
54	130
55	119
56	153

Regression for Time Series: K – Nearest Neighbours: Example

- To implement 5-nn we search in months 1 to 46, the 5 nearest neigbours of 190.
  - Month 10: 190, Month 11: 192
  - Month 11: 192, Month 12: 192
  - Month 12: 192, Month 13: 147
  - Month 24: 188, Month 25: 139
  - Month 36: 184, Month 37: 151

- The 5-NN prediction is then (192+192+147+139+151)/5 = 164.2

10	190	0
11	192	2
12	192	2
13	147	43
14	133	57
15	163	27
16	150	40
17	129	61
18	131	59
19	145	45
20	137	53
21	138	52
22	168	22
23	176	14
24	188	2
25	139	51
26	143	47
27	150	40