36-467 Homework 4

Eu Jing Chua

September 24, 2018

Question 1

Q1 a)

$$Cov[X(\text{DUB}, t), X(\text{DUB}, t-1)] = 14.537$$

Q1 b)

$$\beta = \frac{Cov[X(\text{DUB}, t), X(\text{DUB}, t-1)]}{Var[X(\text{DUB}, t-1)]} \tag{1}$$

$$= 0.58705 \tag{2}$$

$$\alpha = \mathbb{E}[X(\text{DUB}, t)] - \beta \cdot \mathbb{E}[X(\text{DUB}, t-1)] \tag{3}$$

$$= 4.0459 \tag{4}$$

$$Var[X(\text{DUB}, t) - m(X(\text{DUB}, t-1))] = 16.241 \tag{5}$$

Q1 c)

The intercept is 4.0459 and the slope is 0.58705, which is same as the values from the optimal linear predictor.

Q1 d)

The mean of the errors is 0 and the variance is 16.241, which is the same as the values found in Q1 b).

Q1 e)

The intercept will be 4.0472 and the slope is 0.58678.

Question 2

Q2 a)

The intercept will be 6.3837 and the slope is 0.3484.

Q2 b) Let $\vec{Z} = [X(\text{DUB}, t-1), X(\text{DUB}, t-2)]$ and $Y = X(\text{DUB}, t)$

Q2 c)

Table 1: Variance matrix of Z, Var[Z]

	X(DUB, t-1)	X(DUB, t-2)
X(DUB, t-1)	24.76489	14.52755
X(DUB, t-2)	14.52755	24.75673

Table 2: Covariance between Y and Z, Cov[Y, Z]

	X(DUB, t)
X(DUB, t-1)	14.53847
X(DUB, t-2)	8.62537

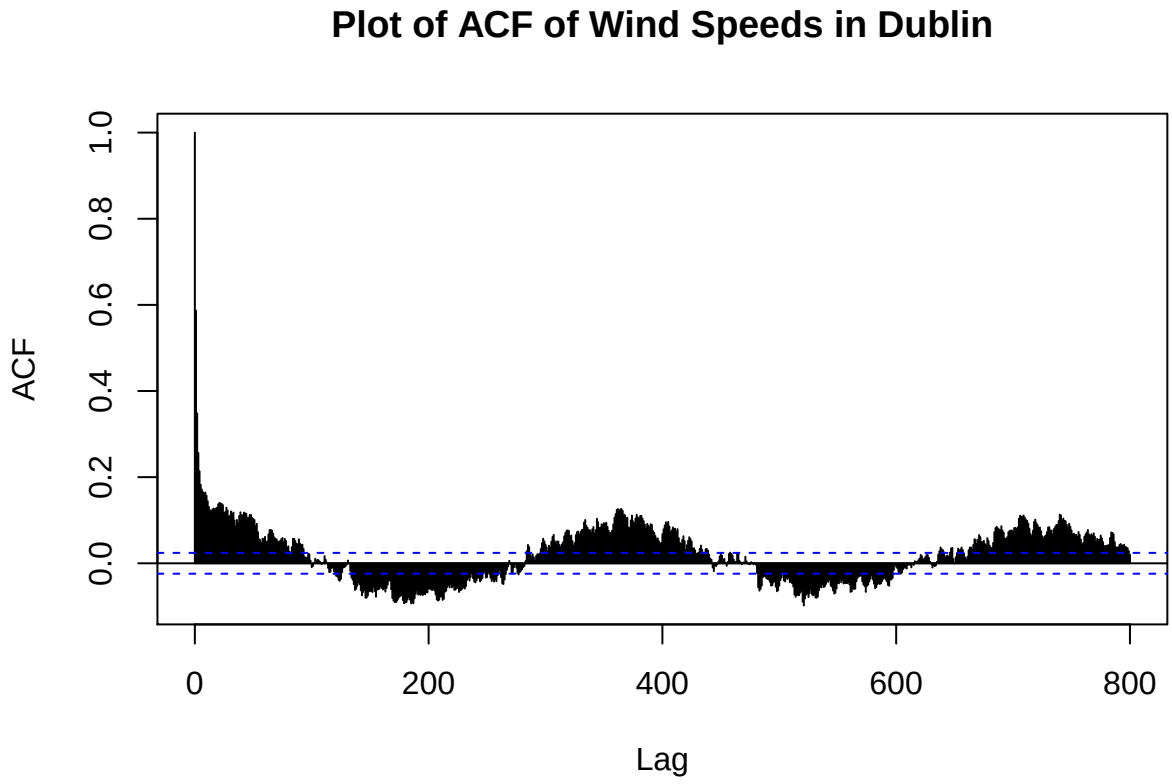
The intercept is 4.0217 and the slopes are as follows:

	Slope
X(DUB, t-1)	0.58356
X(DUB, t-2)	0.00596

Q2 d) The relative difference between the two slopes in the multivariate regression is much larger than that in the two univariate regressions as in the multivariate case, we consider the covariance between $X(\text{DUB}, t-1)$ and $X(\text{DUB}, t-2)$ when determining the slopes. This is on top of just the variances of $X(\text{DUB}, t-1)$ and $X(\text{DUB}, t-2)$ individually, hence the difference in difference between the two slopes.

Question 3

Q3 a)

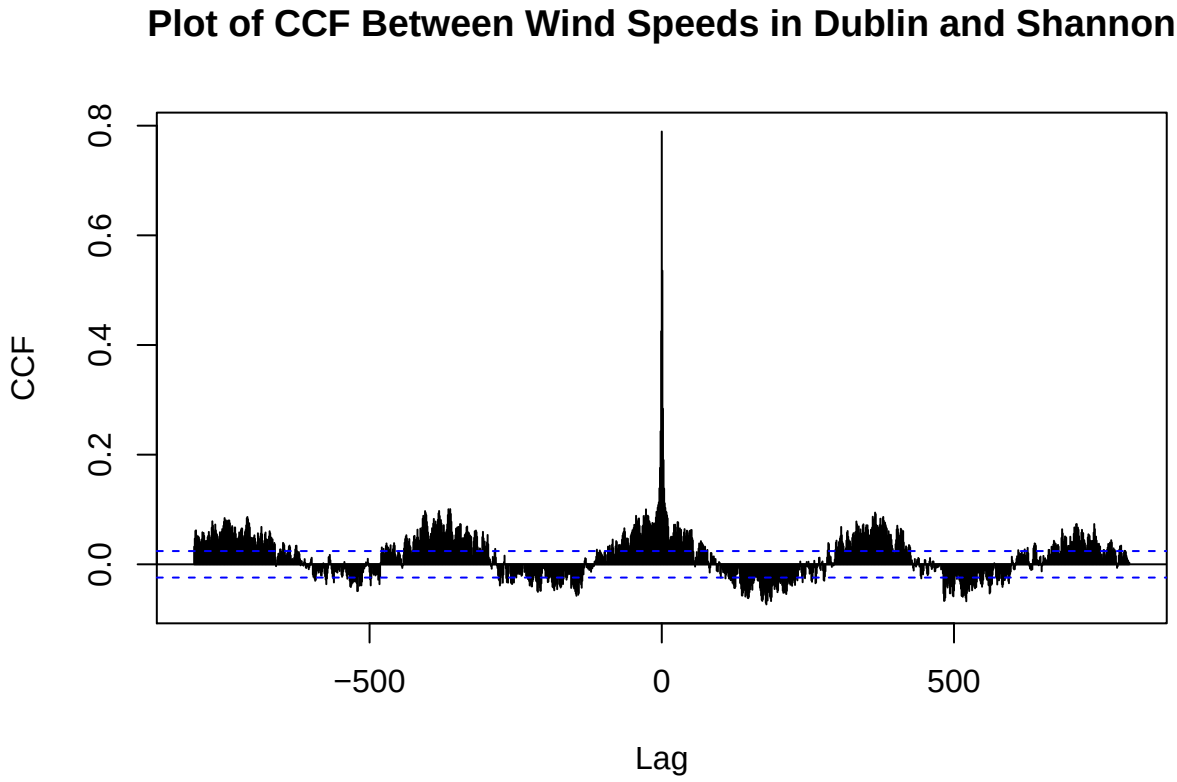


The shape of the function seems to be sinusoidal in nature, shortly after the point $lag = 0$. Looking at the rough position of the peaks and troughs, we can see that the period of oscillation is approximately 365 days (1 year), which would seem to indicate that there is some sort of annual pattern going on.

Q3 b)

The slope is 0.13424 for predicting from $X(\text{DUB}, t-365)$.
As predicting from $X(\text{DUB}, t-1)$ and $X(\text{DUB}, t-365)$, the slopes are as follows:

	Slope
X(DUB, t-1)	0.58442
X(DUB, t-365)	0.05632



Q3 c)

Again, the shape of the function seems to be sinusoidal, except around $lag = 0$. The rough positions of the peaks and troughs also again hint at an annual pattern in the period of oscillation.

Question 4

Q4 a)

Table 3: Variance matrix of wind speeds at locations

	RPT	VAL	ROS	KIL	SHA	BIR	CLA	MUL	CLO	BEL	MAL
RPT	31.580	24.918	20.424	17.541	23.052	18.137	19.264	18.284	19.057	21.315	23.503
VAL	24.918	27.758	15.781	14.718	22.279	16.948	18.987	16.258	17.144	23.054	21.397
ROS	20.424	15.781	25.078	13.437	14.544	12.959	13.658	12.941	13.704	13.750	16.031
KIL	17.541	14.718	13.437	12.999	15.314	12.459	13.244	12.723	13.693	14.650	15.885
SHA	23.052	22.279	14.544	15.314	24.362	17.720	19.388	17.555	18.222	22.162	22.245
BIR	18.137	16.948	12.959	12.459	17.720	15.750	15.858	14.804	15.623	18.152	18.803
CLA	19.264	18.987	13.658	13.244	19.388	15.858	20.241	16.230	17.810	22.676	22.444
MUL	18.284	16.258	12.941	12.723	17.555	14.804	16.230	17.370	16.467	18.776	21.987
CLO	19.057	17.144	13.704	13.693	18.222	15.623	17.810	16.467	20.283	21.290	24.326
BEL	21.315	23.054	13.750	14.650	22.162	18.152	22.676	18.776	21.290	34.048	29.573
MAL	23.503	21.397	16.031	15.885	22.245	18.803	22.444	21.987	24.326	29.573	44.861

Table 4: Cov[X(DUB, t), X(r, t)]

RPT	20.575
VAL	17.492
ROS	16.518
KIL	14.624
SHA	19.391
BIR	16.379
CLA	17.781
MUL	18.261
CLO	18.876
BEL	20.320
MAL	25.588

Q4 b)

Table 5: Coefficients of Optimal Linear Predictor

	Coefficient
Intercept	-0.51604
X(RPT, t)	-0.04197
X(VAL, t)	-0.10438
X(ROS, t)	0.17313
X(KIL, t)	0.10153
X(SHA, t)	0.16231
X(BIR, t)	0.07101
X(CLA, t)	-0.05806
X(MUL, t)	0.51996
X(CLO, t)	0.18208
X(BEL, t)	-0.04778
X(MAL, t)	0.14105

Q4 c)

The theoretical variance of the prediction errors is 4.2696.

Q4 d)

The mean of the prediction errors is 0, while the variance is 4.2696.

Question 5

Q5 a)

Table 6: Variance matrix of wind speeds at locations

	RPT	VAL	ROS	KIL	SHA	BIR	CLA	MUL	CLO	BEL	MAL
RPT	31.575	24.914	20.408	17.539	23.054	18.136	19.263	18.283	19.056	21.319	23.499
VAL	24.914	27.755	15.768	14.717	22.280	16.947	18.986	16.258	17.143	23.058	21.393
ROS	20.408	15.768	25.044	13.431	14.542	12.954	13.652	12.936	13.700	13.754	16.018
KIL	17.539	14.717	13.431	12.999	15.315	12.460	13.245	12.723	13.694	14.653	15.884
SHA	23.054	22.280	14.542	15.315	24.365	17.722	19.391	17.557	18.225	22.166	22.247
BIR	18.136	16.947	12.954	12.460	17.722	15.751	15.859	14.805	15.624	18.156	18.803
CLA	19.263	18.986	13.652	13.245	19.391	15.859	20.243	16.231	17.812	22.679	22.444
MUL	18.283	16.258	12.936	12.723	17.557	14.805	16.231	17.372	16.469	18.779	21.988
CLO	19.056	17.143	13.700	13.694	18.225	15.624	17.812	16.469	20.285	21.293	24.327
BEL	21.319	23.058	13.754	14.653	22.166	18.156	22.679	18.779	21.293	34.053	29.578
MAL	23.499	21.393	16.018	15.884	22.247	18.803	22.444	21.988	24.327	29.578	44.862

Table 7: Cov[X(DUB, t), X(r, t-1)]

RPT	13.815
VAL	13.313
ROS	9.393
KIL	8.521
SHA	13.161
BIR	10.588
CLA	12.025
MUL	10.920
CLO	11.607
BEL	14.979
MAL	16.316

Q5 b)

Table 8: Coefficients of Optimal Linear Predictor

	Coefficient
Intercept	2.42612
X(RPT, t-1)	0.10154
X(VAL, t-1)	0.05548
X(ROS, t-1)	0.06715
X(KIL, t-1)	-0.30568
X(SHA, t-1)	0.18003
X(BIR, t-1)	0.12404
X(CLA, t-1)	0.02828
X(MUL, t-1)	0.07055
X(CLO, t-1)	0.03548
X(BEL, t-1)	0.09766
X(MAL, t-1)	0.09468

When comparing the current slopes to the slopes found in 4 b), it seems that some of the slopes (RPT, VAL, SHA, BIR, CLA, BEL) increased in value, whereas the rest decreased.

Q5 c)

The theoretical variance of the prediction errors is 16.395.

Q5 d)

The mean of the prediction errors is 0, while the variance is 16.395.

Question 6

Q6 a)

Table 9: Root-mean-square of Prediction Errors

	RMSE
1b)	4.02967
2c)	4.02990
4b)	2.06615
5b)	4.04874

It seems that the predictor in the model from 4b) does the best as it has the lowest root-mean-square error.

Q6 b)

Table 10: Coefficients of OLP in 1b)

	Old	New
Intercept	4.04590	4.34849
X(DUB, t-1)	0.58705	0.57156

Table 11: Coefficients of OLP in 2c)

	Old	New
Intercept	4.02171	4.31648
X(DUB, t-1)	0.58356	0.56739
X(DUB, t-2)	0.00596	0.00734

Table 12: Coefficients of OLP in 4b)

	Old	New
Intercept	-0.51604	-0.53200
X(RPT, t)	-0.04197	-0.02873
X(VAL, t)	-0.10438	-0.05713
X(ROS, t)	0.17313	0.20091
X(KIL, t)	0.10153	-0.02817
X(SHA, t)	0.16231	0.12155
X(BIR, t)	0.07101	0.05318
X(CLA, t)	-0.05806	-0.12654
X(MUL, t)	0.51996	0.72049
X(CLO, t)	0.18208	0.17931
X(BEL, t)	-0.04778	-0.04187
X(MAL, t)	0.14105	0.11980

Table 13: Coefficients of OLP in 5b)

	Old	New
Intercept	2.42612	2.63484
X(RPT, t-1)	0.10154	0.14101
X(VAL, t-1)	0.05548	0.03750
X(ROS, t-1)	0.06715	0.07705
X(KIL, t-1)	-0.30568	-0.36670
X(SHA, t-1)	0.18003	0.11891
X(BIR, t-1)	0.12404	0.22284
X(CLA, t-1)	0.02828	-0.04879
X(MUL, t-1)	0.07055	0.12258
X(CLO, t-1)	0.03548	0.02714
X(BEL, t-1)	0.09766	0.12458
X(MAL, t-1)	0.09468	0.09915

Q6 c)

Table 14: Predictions of Jan 1 1970 Wind Speeds

	Predictions
1b)	12.68178
2c)	12.71229
4b)	8.82371
5b)	12.70167

Table 15: Prediction Errors

	Error
1b)	-3.68178
2c)	-3.71229
4b)	0.17629
5b)	-3.70167

Q6 d)

Table 16: Mean Prediction Errors

	Mean Prediction Errors
1b)	-0.30168
2c)	-0.29965
4b)	-0.89523
5b)	-0.86979

Table 17: Root-mean-square Prediction Errors

	RMSE
1b)	3.85043
2c)	3.85031
4b)	2.27741
5b)	3.94911

Q6 e) The numbers differ as in 6a), the models were fitted using the entire data set and then used to predict values it used for fitting, while in 6d) the models were fitted using only a portion of the data set and then the held out portion was used for calculating prediction error. The prediction errors in 6d) give a better idea as it gives us an idea of how the models we have created perform on data it has not seen before, therefore giving us a better sense of generalization.