Homework 2

Deanna Springgay

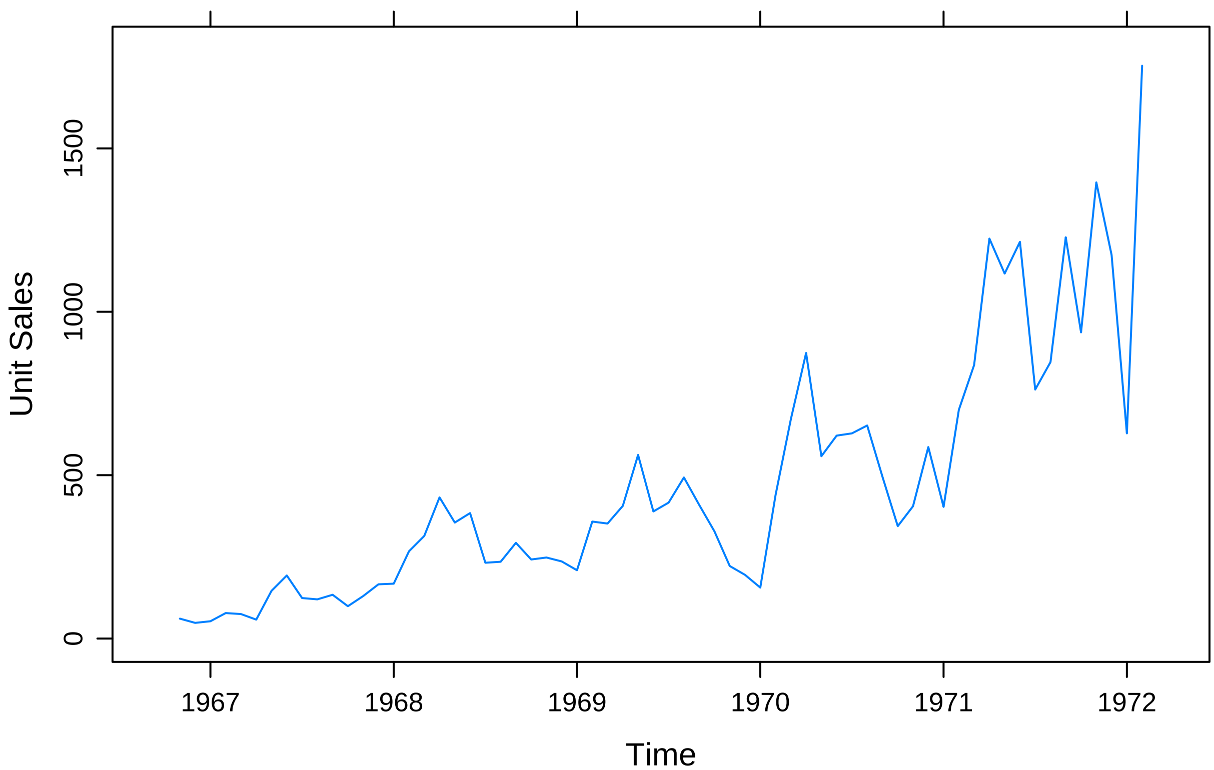
2/19/2021

Problems

3.2 and 3.3 are attached at the end

3.7a)

From the following plot we can see that the Unit Sales has been following an increasing trend over the time period of 1967 to 1972.



3.7b)

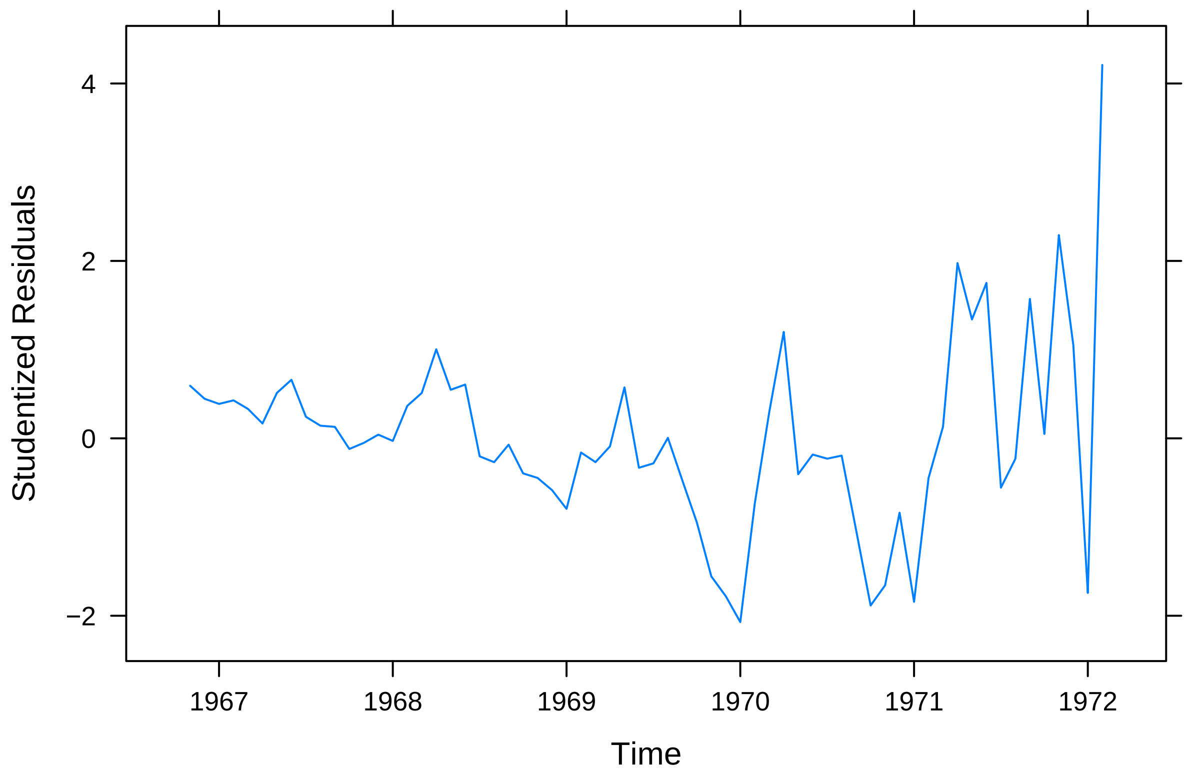
Analysis after fitting a linear model indicates that both the intercept and slope of the linear model are significantly significant since the p-values are so small at a scale of e-17. 69% of the variance present in the data is explained by the model.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-394886	33540	-11.77	1.87e-17
time(winnebago)	200.7	17.03	11.79	1.777e-17

Table 2: Fitting linear model: winnebago ~ time(winnebago)

Observations	Residual Std. Error	R^2	Adjusted R^2
64	209.7	0.6915	0.6865

The following plot of the residuals shows the range of the residuals increasing as time passes, indicating heteroscedasticity and that this data will need to be transformed before we continue analysis.



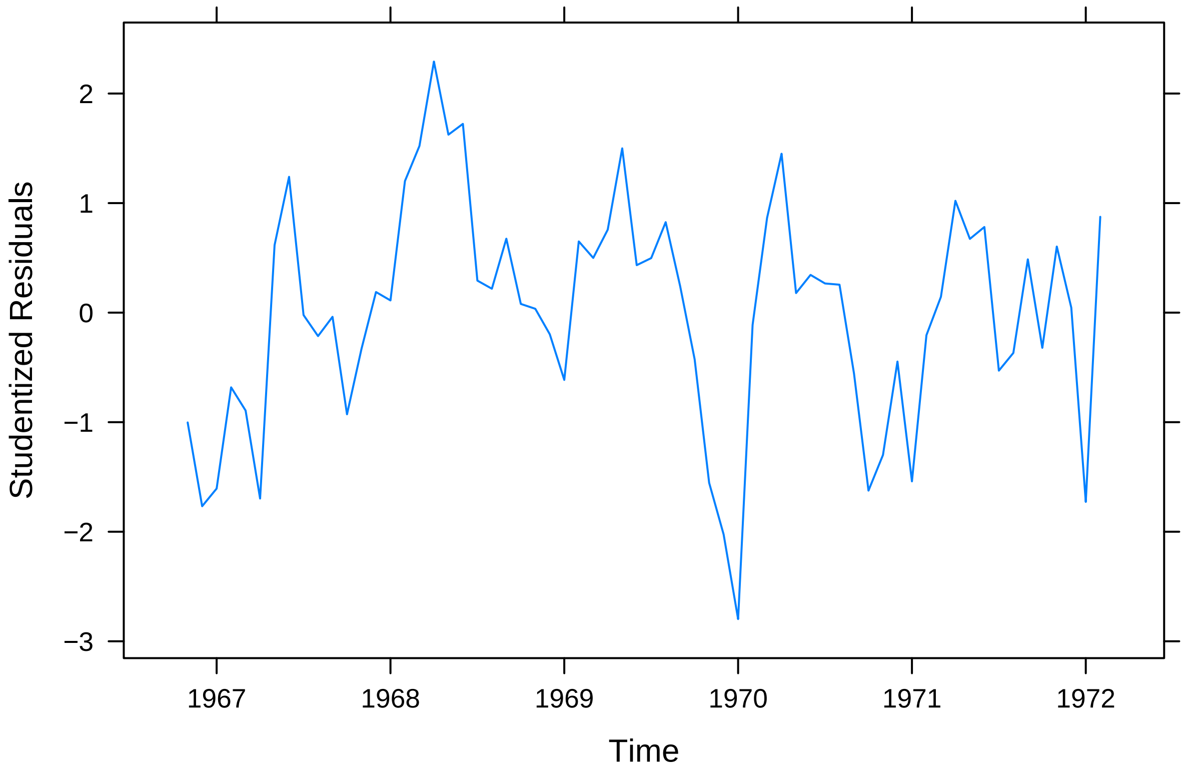
3.7c)

The following plot shows the data after undergoing a log transformation, and it is apparent that the increasing variance has been corrected. The residuals now appear to have a zero mean and constant variance.This model is better, as almost 80% of the variance is explained by the model.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-984.9	62.99	-15.64	3.45e-23
time(winnebago)	0.5031	0.03199	15.73	2.575e-23

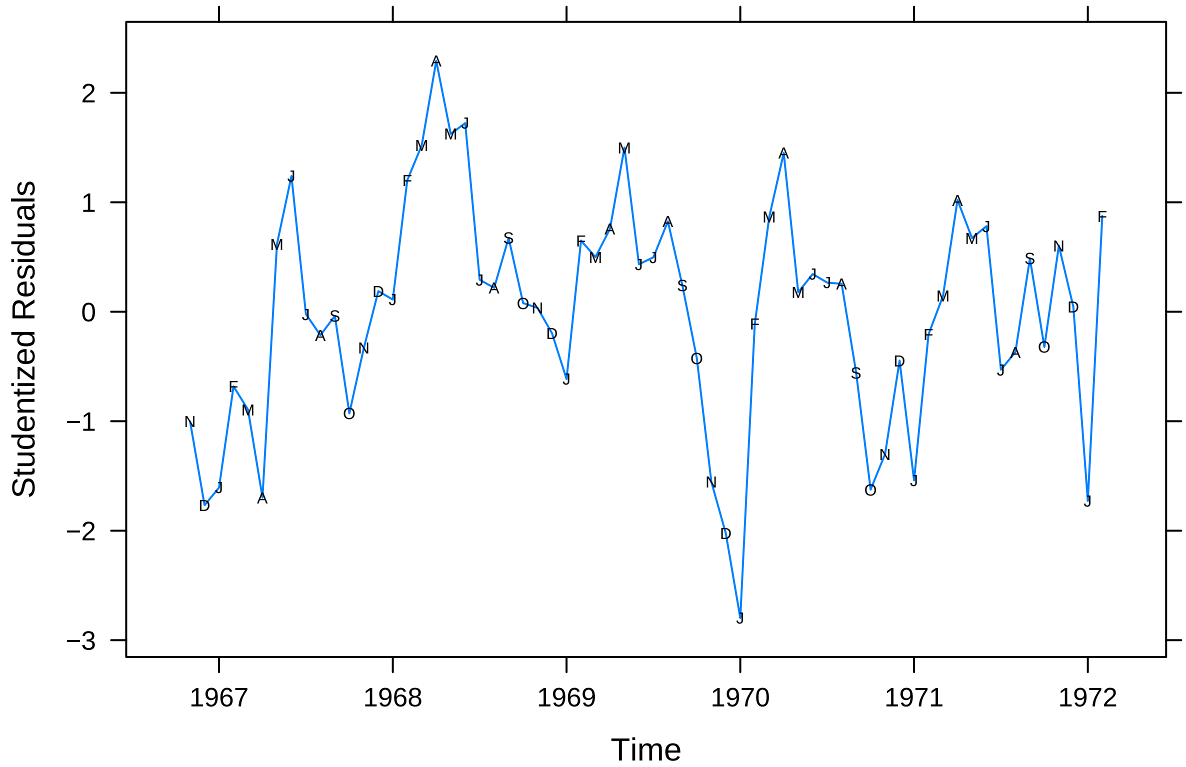
Table 4: Fitting linear model: log(winnebago) ~ time(winnebago)

Observations	Residual Std. Error	R^2	Adjusted R^2
64	0.3939	0.7996	0.7964



3.7d)

From the following plot we can see a somewhat oscillating pattern of higher residuals around springtime in comparison to the other months. Overall, the residuals look more like random noise, though it is apparent that we are overestimating values for the springtime.



3.7e)

This model further explains variance found in the data; it explains 89% of the variance. The following table summarizes the significant seasonal-means and time trends:

Significant	Not Significant
Intercept	October
February	November
March	December
April	
May	
June	
July	
August	
September	
Time	

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-997.3	50.64	-19.69	1.718e-25
season(winnebago)February	0.6244	0.1818	3.434	0.001188
season(winnebago)March	0.6822	0.1909	3.574	0.0007793
season(winnebago)April	0.8096	0.1908	4.243	9.301e-05
season(winnebago)May	0.8695	0.1907	4.559	3.246e-05
season(winnebago)June	0.8631	0.1907	4.526	3.627e-05
season(winnebago)July	0.5539	0.1907	2.905	0.00542

	Estimate	Std. Error	t value	Pr(>|t|)
season(winnebago)August	0.5699	0.1907	2.988	0.004305
season(winnebago)September	0.5757	0.1907	3.018	0.00396
season(winnebago)October	0.2635	0.1908	1.381	0.1733
season(winnebago)November	0.2868	0.1819	1.577	0.1209
season(winnebago)December	0.248	0.1818	1.364	0.1785
time(winnebago)	0.5091	0.02571	19.8	1.351e-25

Table 7: Fitting linear model: log(winnebago) ~ season(winnebago) + time(winnebago)

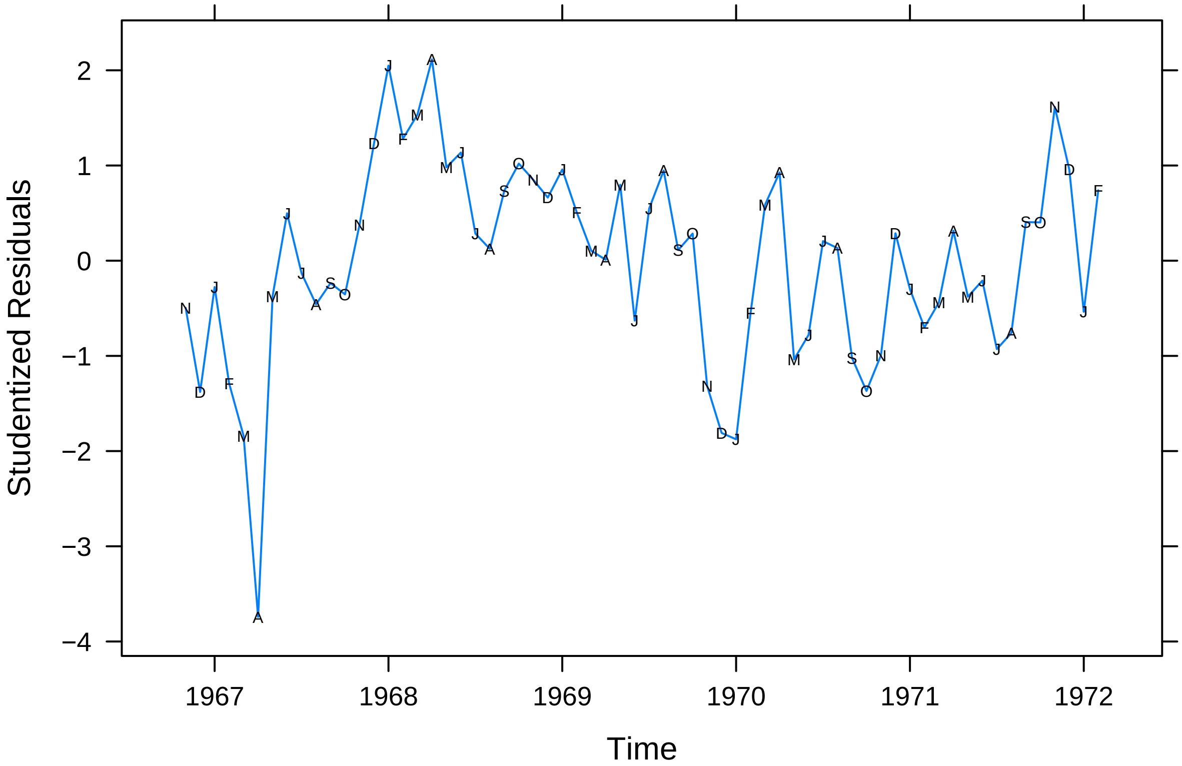
Observations	Residual Std. Error	R^2	Adjusted R^2
64	0.3149	0.8946	0.8699

The following HAC test indicates that only October and November are not influenced by other values.

```
##
## t test of coefficients:
##
##                               Estimate  Std. Error  t value  Pr(>|t|)
## (Intercept)                -997.330613   98.209008 -10.1552 7.610e-14 ***
## season(winnebago)February     0.624448    0.123094   5.0729 5.547e-06 ***
## season(winnebago)March        0.682197    0.175580   3.8854 0.0002957 ***
## season(winnebago)April        0.809588    0.264416   3.0618 0.0035066 **
## season(winnebago)May          0.869525    0.133102   6.5328 2.977e-08 ***
## season(winnebago)June         0.863087    0.154476   5.5872 9.022e-07 ***
## season(winnebago)July         0.553918    0.170223   3.2541 0.0020224 **
## season(winnebago)August       0.569885    0.191367   2.9780 0.0044321 **
## season(winnebago)September    0.575717    0.174867   3.2923 0.0018089 **
## season(winnebago)October      0.263486    0.196040   1.3440 0.1848837
## season(winnebago)November     0.286822    0.163248   1.7570 0.0849241 .
## season(winnebago)December     0.248022    0.113394   2.1873 0.0333330 *
## time(winnebago)               0.509090    0.049835  10.2155 6.204e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

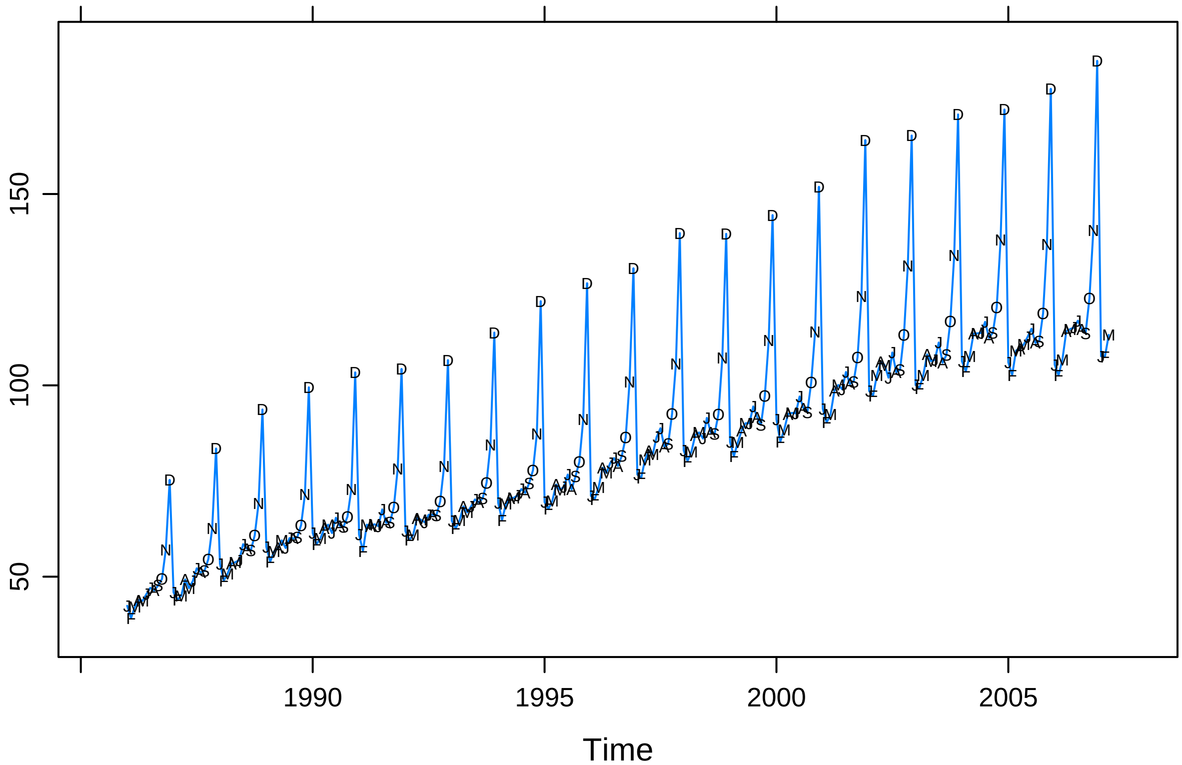
3.7f)

The following residual plot indicates that the residuals follow a zero mean and constant variance.



3.8a)

From this plot it appears that retail sales in the UK have been rising each year, but with a large peak each year in December, likely due to holiday shopping.



3.8b)

From these residuals we can conclude that every month, intercept, and time are significantly significant except for the month of March. This model also appears to be an effective model; accounting for almost 98% of the variance in the data.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-7249	87.24	-83.1	6.41e-180
season(retail)February	-3.015	1.29	-2.337	0.02024
season(retail)March	0.07469	1.29	0.05791	0.9539
season(retail)April	3.447	1.305	2.641	0.008801
season(retail)May	3.108	1.305	2.381	0.01803
season(retail)June	3.074	1.305	2.355	0.01932
season(retail)July	6.053	1.305	4.638	5.757e-06
season(retail)August	3.138	1.305	2.404	0.01695
season(retail)September	3.428	1.305	2.626	0.009187
season(retail)October	8.555	1.305	6.555	3.336e-10
season(retail)November	20.82	1.305	15.95	1.274e-39
season(retail)December	52.54	1.305	40.25	3.169e-109
time(retail)	3.67	0.04369	84	5.206e-181

Table 9: Fitting linear model: retail ~ season(retail) + time(retail)

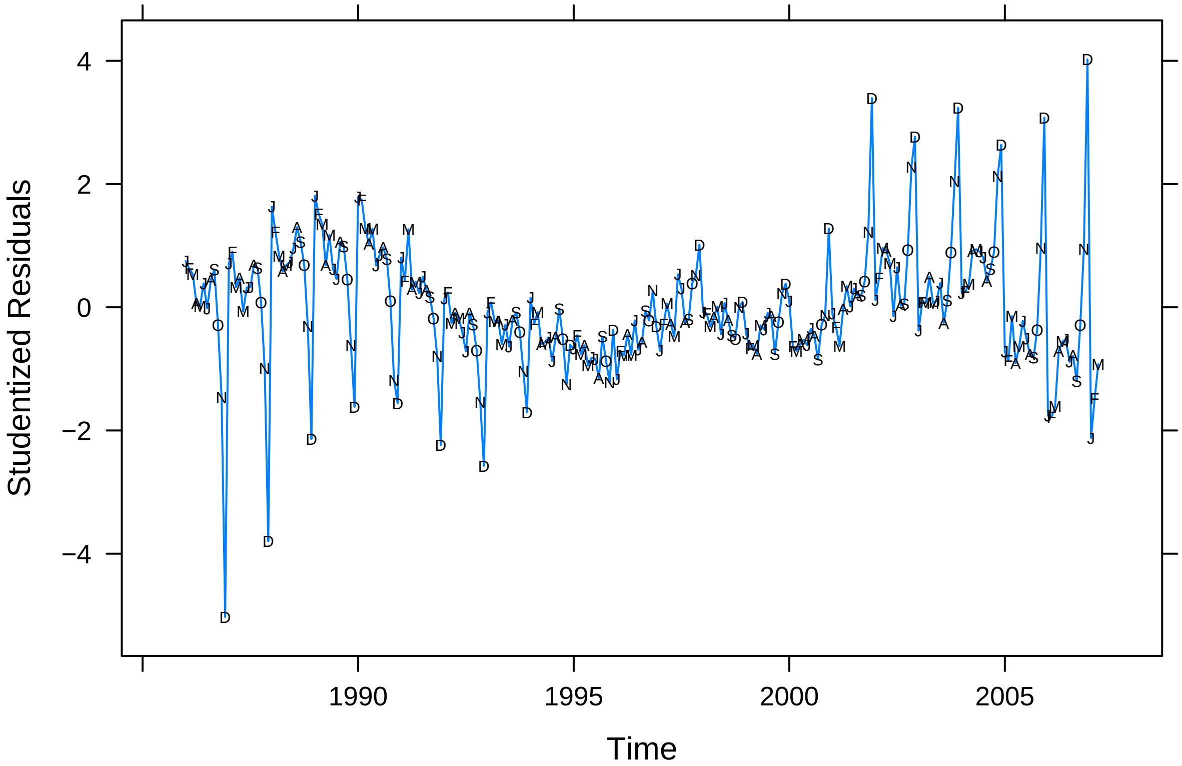
Observations	Residual Std. Error	R^2	Adjusted R^2
255	4.278	0.9767	0.9755

The following HAC test indicates that only March is not influenced by other values.

```
##
## t test of coefficients:
##
##                            Estimate  Std. Error  t value  Pr(>|t|)
## (Intercept)              -7.2494e+03  1.3584e+02 -53.3652 < 2.2e-16 ***
## season(retail)February  -3.0149e+00  4.2128e-01  -7.1565 9.825e-12 ***
## season(retail)March      7.4693e-02  6.8990e-01   0.1083 0.9138740
## season(retail)April      3.4472e+00  8.9649e-01   3.8452 0.0001541 ***
## season(retail)May        3.1080e+00  9.7418e-01   3.1903 0.0016088 **
## season(retail)June       3.0736e+00  9.8755e-01   3.1123 0.0020792 **
## season(retail)July       6.0535e+00  1.0472e+00   5.7808 2.282e-08 ***
## season(retail)August     3.1381e+00  1.0185e+00   3.0812 0.0022998 **
## season(retail)September  3.4275e+00  9.4676e-01   3.6203 0.0003584 ***
## season(retail)October    8.5550e+00  9.4500e-01   9.0530 < 2.2e-16 ***
## season(retail)November   2.0816e+01  1.6325e+00  12.7506 < 2.2e-16 ***
## season(retail)December   5.2543e+01  2.9377e+00  17.8859 < 2.2e-16 ***
## time(retail)             3.6700e+00  6.8009e-02  53.9637 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

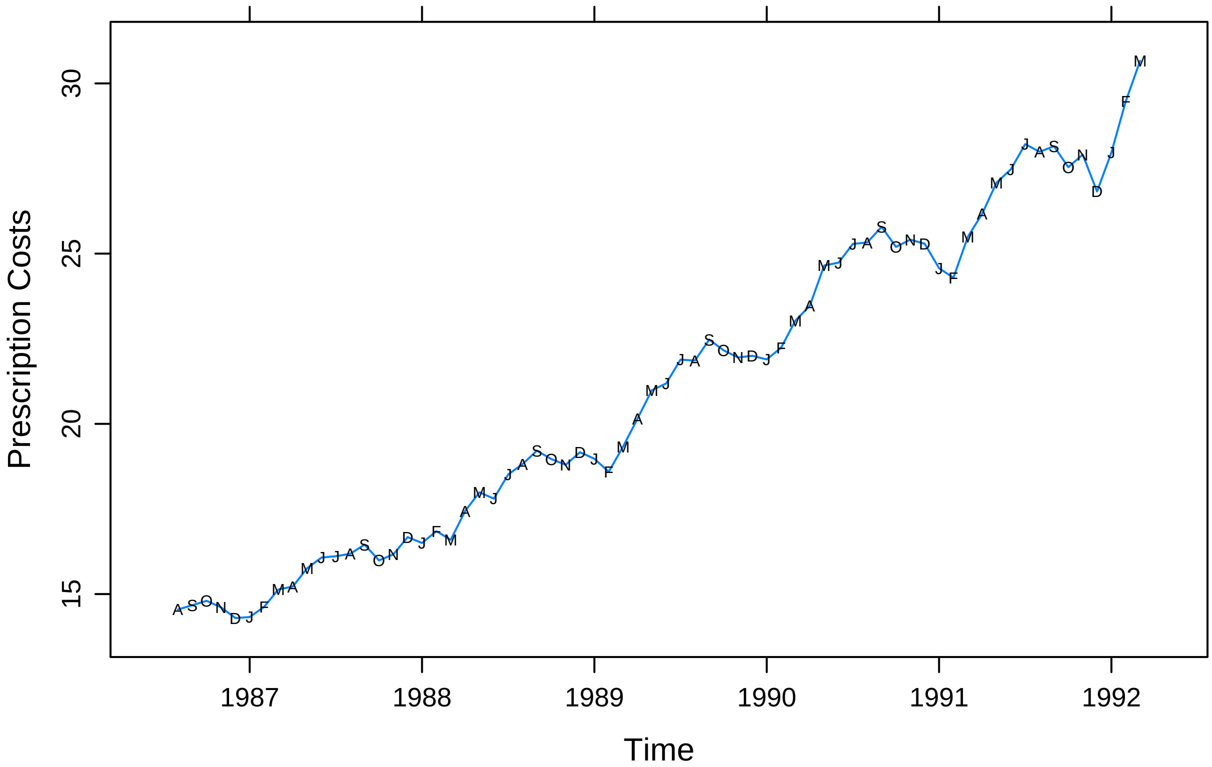
3.8c)

While the residuals from this time period seem to average a zero mean and roughly have constant variance, there appears to be somewhat of a pattern when it comes to December: the residuals seem to follow a positive linear trend while the other months stay around zero. This makes me hesitate in claiming constant variance, since the second half of the time period could be described as showing heteroscedasticity.



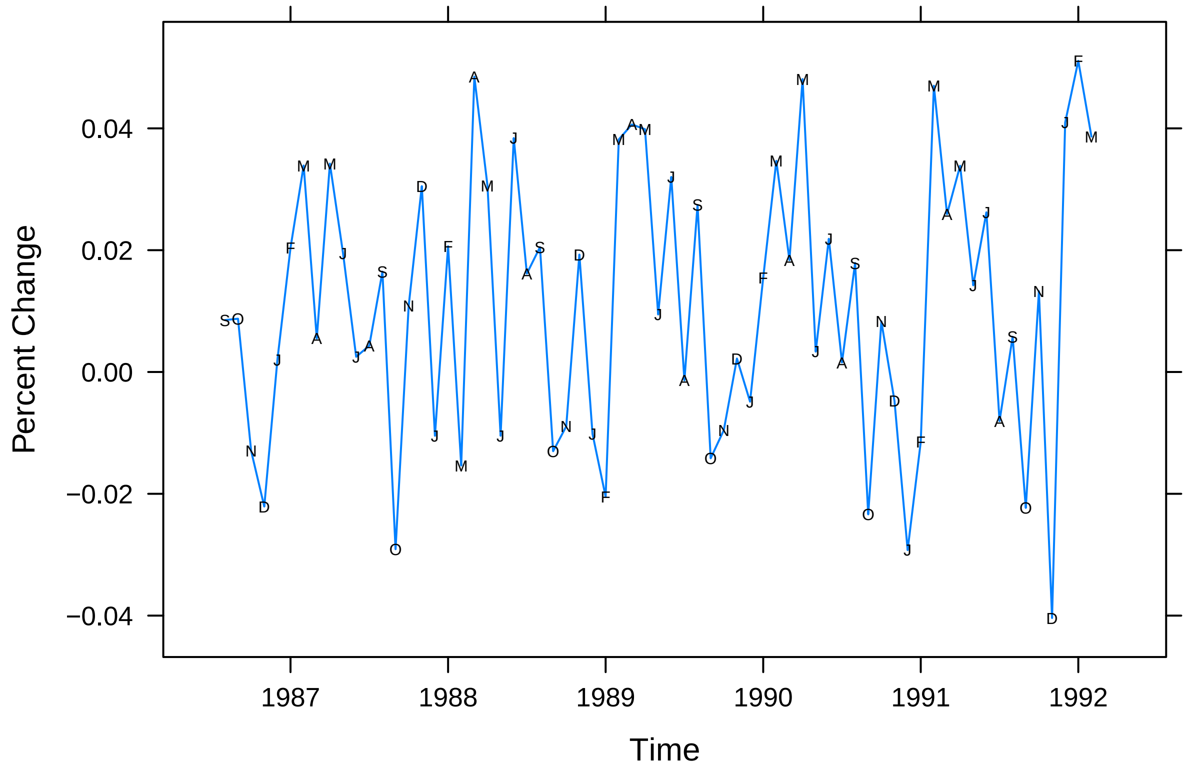
3.9a)

From this plot we can see a cyclical seasonal trend with the peak season being summer and there is an increase over the long-term.



3.9b)

The plot shows that the monthly percentage change appears to be stationary.



3.9c)

While the model is statistically significant (all portions of the model have p-values less than 0.05), the model only explains about 31% of the variance present in the data.

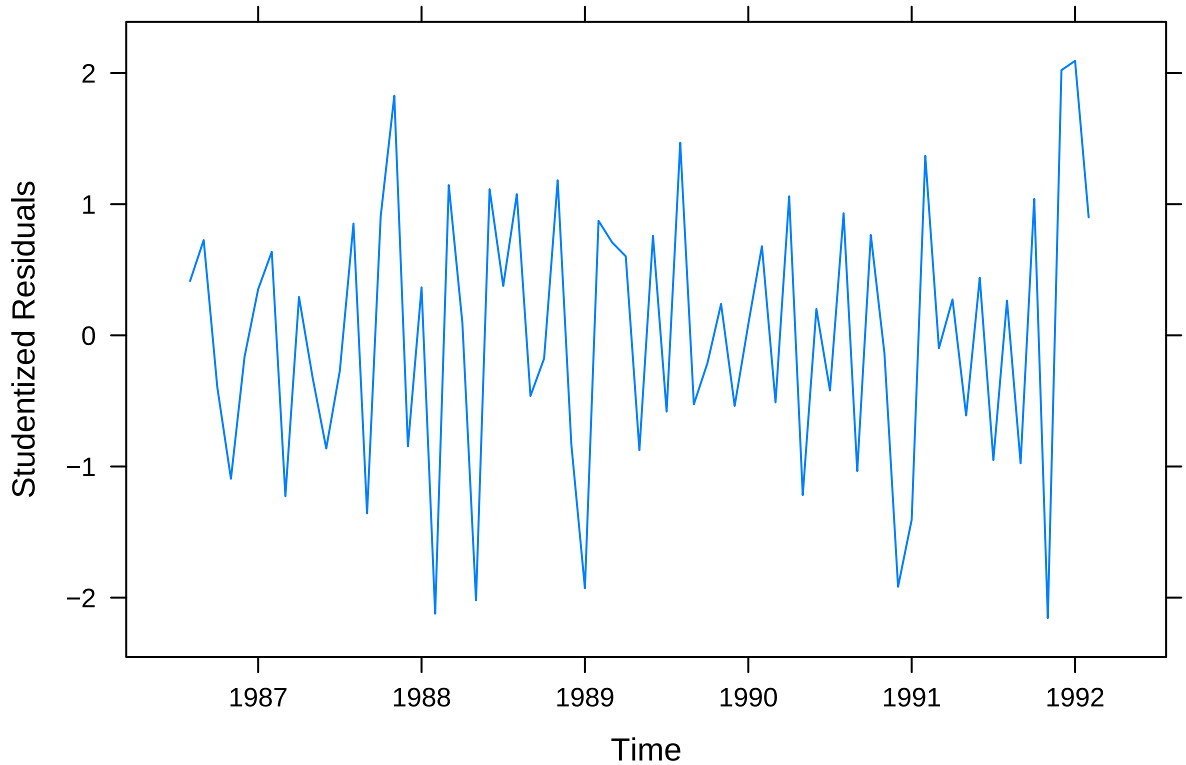
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.01159	0.002282	5.08	3.508e-06
harmonic(pchange)cos(2*pi*t)	-0.006605	0.003237	-2.041	0.04542
harmonic(pchange)sin(2*pi*t)	0.01612	0.003208	5.026	4.291e-06

Table 11: Fitting linear model: pchange ~ harmonic(pchange)

Observations	Residual Std. Error	R^2	Adjusted R^2
67	0.01862	0.3126	0.2912

3.9d)

The following residual plot appears to have a zero mean and constant variance, indicating that the coside trend model is adequate.



3.13a)

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-3.742	-0.5678	0.1178	-0.007572	0.7369	2.115

Table 13: Table continues below

1	2	3	4	5	6	7	8
-0.4949	-1.379	-0.2769	-1.289	-1.84	-3.742	-0.3731	0.4978

Table 14: Table continues below

9	10	11	12	13	14	15	16	17
-0.1315	-0.4585	-0.2356	-0.3536	0.3755	1.23	2.049	1.281	1.534

Table 15: Table continues below

18	19	20	21	22	23	24	25	26
2.115	0.9808	1.136	0.2847	0.1237	0.736	1.02	0.8497	0.663

Table 16: Table continues below

27	28	29	30	31	32	33	34
0.9588	0.5077	0.1048	0.009551	0.7978	-0.6296	0.5481	0.9466

Table 17: Table continues below

35	36	37	38	39	40	41	42	43
0.1119	0.2847	-1.316	-1.81	-1.876	-0.5472	0.5875	0.9269	-1.037

Table 18: Table continues below

44	45	46	47	48	49	50	51
-0.7804	0.2056	0.1318	-1.022	-1.368	-0.9935	0.2879	-0.2992

Table 19: Table continues below

52	53	54	55	56	57	58	59
-0.7012	-0.4377	0.3152	-0.3787	-0.2092	-0.9285	-0.7598	0.405

60	61	62	63	64
0.4026	1.614	0.9576	-0.5369	0.7396

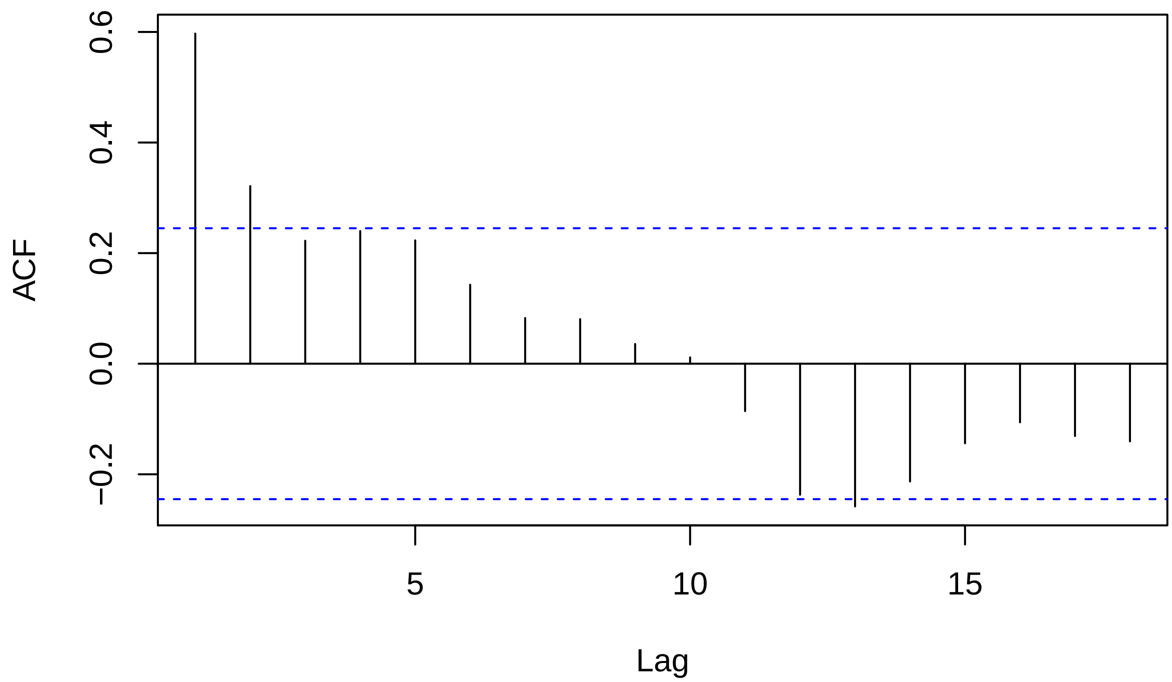
3.13b)

The runs test is significant since we have less runs than expected. We can conclude that the residuals are not random.

```
## $pvalue
## [1] 0.000243
##
## $observed.runs
## [1] 18
##
## $expected.runs
## [1] 32.71875
##
## $n1
## [1] 29
##
## $n2
## [1] 35
##
## $k
## [1] 0
```

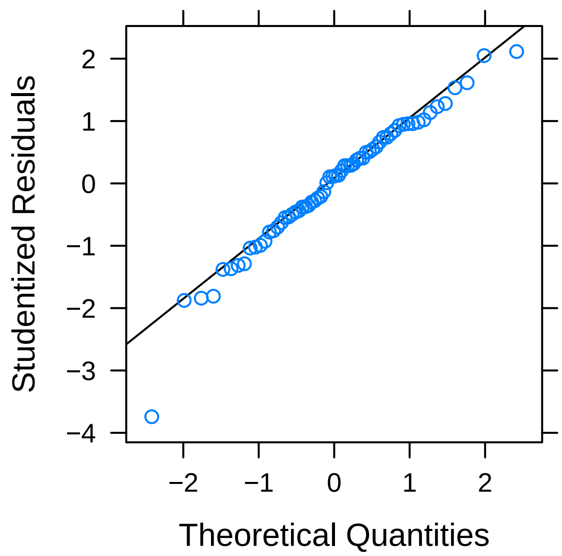
3.13c)

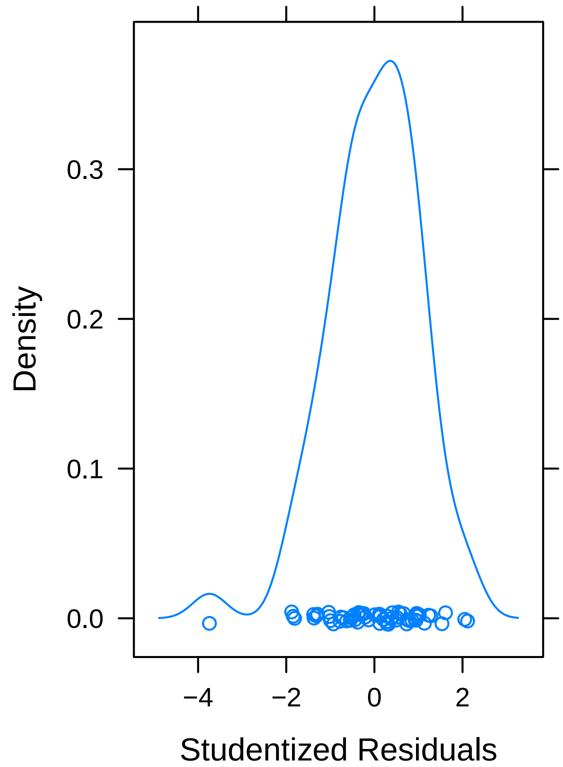
The following ACF plot indicates there is autocorrelation present in the residuals, so we can conclude there is dependency present.



3.13d)

The Q-Q plot indicates there is a left-skew in the residuals. The density plot shows us an approximately normal distribution among the residuals.





3.14a)

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-5.026	-0.5848	-0.08414	0.0003105	0.5268	4.023

Table 22: Table continues below

1	2	3	4	5	6	7	8
0.7485	0.6303	0.5351	0.05958	0.01952	0.387	-0.01825	0.4409

Table 23: Table continues below

9	10	11	12	13	14	15	16
0.6106	-0.2873	-1.463	-5.026	0.7068	0.9017	0.3255	0.4751

Table 24: Table continues below

17	18	19	20	21	22	23	24
-0.06934	0.3217	0.3252	0.6882	0.6412	0.08043	-0.994	-3.798

Table 25: Table continues below

25	26	27	28	29	30	31	32	33
1.632	1.222	0.8369	0.578	0.6822	0.7372	0.9575	1.298	1.057

Table 26: Table continues below

34	35	36	37	38	39	40	41	42
0.6879	-0.3123	-2.139	1.809	1.519	1.35	0.6808	1.171	0.6236

Table 27: Table continues below

43	44	45	46	47	48	49	50	51
0.4589	1.063	0.9915	0.4544	-0.6169	-1.616	1.79	1.744	1.284

Table 28: Table continues below

52	53	54	55	56	57	58	59	60
1.024	1.274	0.6784	0.8501	0.9727	0.7813	0.1018	-1.187	-1.559

Table 29: Table continues below

61	62	63	64	65	66	67	68	69
0.8052	0.4238	1.266	0.2867	0.4145	0.2297	0.4967	0.2833	0.1647

Table 30: Table continues below

70	71	72	73	74	75	76	77
-0.1784	-0.7937	-2.236	0.1416	0.2394	-0.2615	-0.08907	-0.1768

Table 31: Table continues below

78	79	80	81	82	83	84	85
-0.4093	-0.7175	-0.09248	-0.2828	-0.6981	-1.534	-2.574	-0.09023

Table 32: Table continues below

86	87	88	89	90	91	92	93
0.07919	-0.2303	-0.2254	-0.6005	-0.2823	-0.6381	-0.2049	-0.08414

Table 33: Table continues below

94	95	96	97	98	99	100	101
-0.4031	-1.043	-1.705	0.1558	-0.272	-0.07978	-0.6011	-0.5692

Table 34: Table continues below

102	103	104	105	106	107	108	109
-0.4905	-0.8707	-0.4848	-0.02913	-0.5155	-1.253	-0.6114	-0.6735

Table 35: Table continues below

110	111	112	113	114	115	116	117
-0.4559	-0.766	-0.6178	-0.9457	-0.8188	-0.8393	-1.15	-0.4764

Table 36: Table continues below

118	119	120	121	122	123	124	125
-0.8679	-1.221	-0.3647	-1.17	-0.7119	-0.7827	-0.443	-0.7704

Table 37: Table continues below

126	127	128	129	130	131	132	133
-0.2129	-0.6882	-0.5661	-0.0626	-0.214	0.2721	-0.3097	-0.7069

Table 38: Table continues below

134	135	136	137	138	139	140	141
-0.2743	0.0611	-0.2683	-0.4757	0.5357	0.2998	-0.2478	-0.1989

Table 39: Table continues below

142	143	144	145	146	147	148	149
0.3909	0.5185	1.014	-0.07827	-0.09998	-0.3139	-0.1656	0.009779

Table 40: Table continues below

150	151	152	153	154	155	156
-0.4378	0.06784	-0.2168	-0.4549	-0.5108	-0.0004676	0.0872

Table 41: Table continues below

157	158	159	160	161	162	163	164
-0.4295	-0.6667	-0.6177	-0.7572	-0.294	-0.3589	-0.09238	-0.1379

Table 42: Table continues below

165	166	167	168	169	170	171	172
-0.7595	-0.2404	0.222	0.3815	0.1033	-0.6358	-0.7065	-0.6063

Table 43: Table continues below

173	174	175	176	177	178	179	180
-0.5264	-0.6155	-0.3485	-0.466	-0.8487	-0.2812	-0.1297	1.278

Table 44: Table continues below

181	182	183	184	185	186	187	188
-0.1047	-0.3177	-0.6278	-0.02454	0.3426	0.01422	0.3049	0.1875

Table 45: Table continues below

189	190	191	192	193	194	195	196	197
0.1886	0.4201	1.222	3.395	0.1177	0.4788	0.9598	0.9183	0.7099

Table 46: Table continues below

198	199	200	201	202	203	204	205	206
-0.1463	0.6481	0.02711	0.0522	0.932	2.28	2.77	-0.3777	0.07926

Table 47: Table continues below

207	208	209	210	211	212	213	214
0.08035	0.4934	0.06969	0.1005	0.3917	-0.2533	0.1074	0.8918

Table 48: Table continues below

215	216	217	218	219	220	221	222	223
2.042	3.235	0.228	0.2542	0.3751	0.9101	0.9422	0.9009	0.8079

Table 49: Table continues below

224	225	226	227	228	229	230	231
0.4255	0.6189	0.8997	2.125	2.639	-0.7242	-0.8664	-0.1449

Table 50: Table continues below

232	233	234	235	236	237	238	239
-0.9098	-0.6366	-0.2211	-0.506	-0.7685	-0.8156	-0.3664	0.9641

Table 51: Table continues below

240	241	242	243	244	245	246	247
3.078	-1.76	-1.782	-1.61	-0.7104	-0.5578	-0.5268	-0.8851

248	249	250	251	252	253	254	255
-0.7862	-1.196	-0.2874	0.9482	4.023	-2.123	-1.484	-0.9251

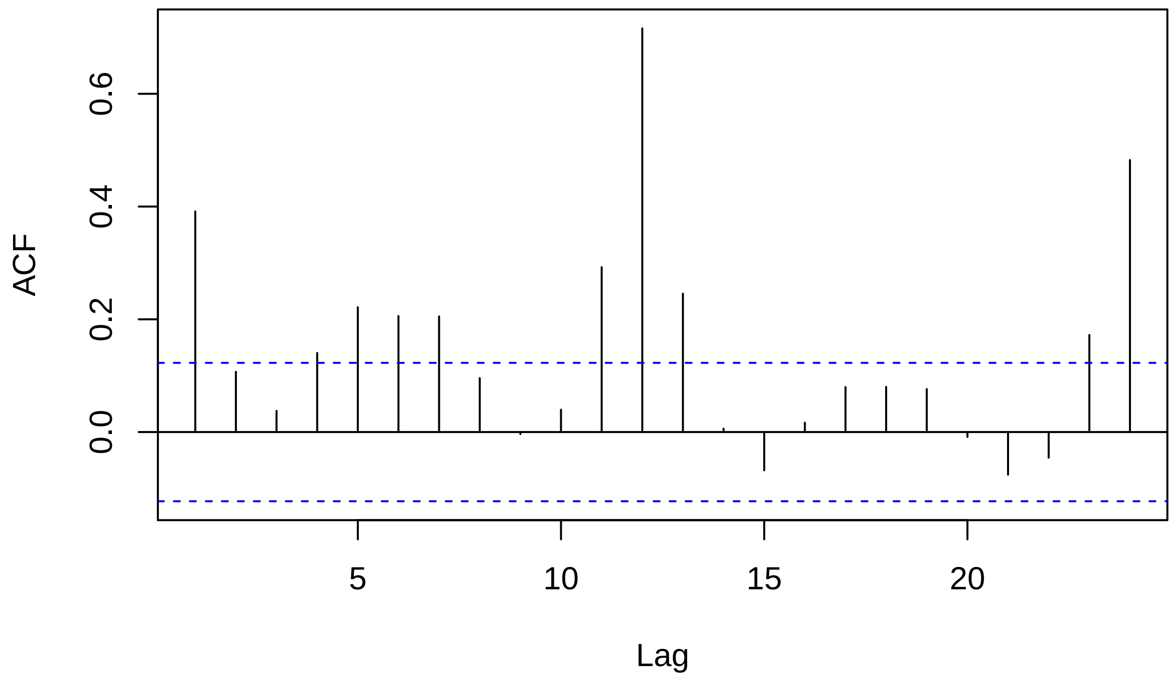
3.14b)

The runs test is significant since we have less runs than expected. We can conclude that the residuals are not random.

```
## $pvalue
## [1] 9.19e-23
##
## $observed.runs
## [1] 52
##
## $expected.runs
## [1] 127.9333
##
## $n1
## [1] 136
##
## $n2
## [1] 119
##
## $k
## [1] 0
```

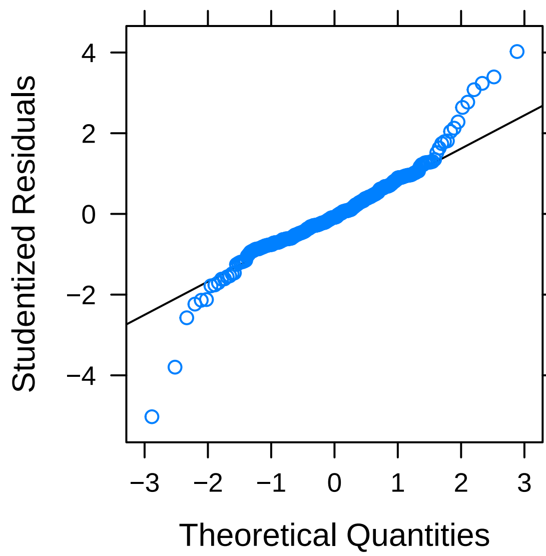
3.14c)

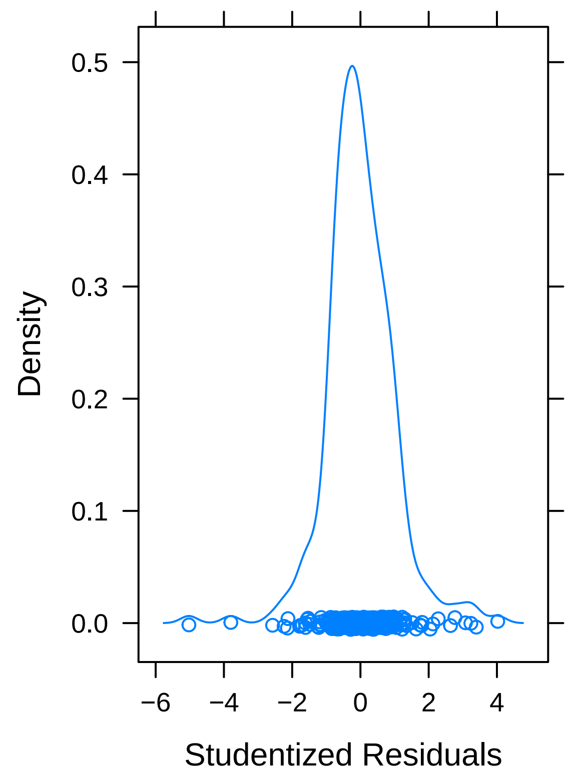
The following ACF plot indicates there is autocorrelation present in the residuals, so we can conclude there is dependency present.



3.14d)

The distribution of the residuals in both the Q-Q plot and the density plot is quite light-tailed.





3.15a)

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.01159	0.002282	5.08	3.508e-06
harmonic(pchange)cos(2*pi*t)	-0.006605	0.003237	-2.041	0.04542
harmonic(pchange)sin(2*pi*t)	0.01612	0.003208	5.026	4.291e-06

Table 54: Fitting linear model: pchange ~ harmonic(pchange)

Observations	Residual Std. Error	R^2	Adjusted R^2
67	0.01862	0.3126	0.2912

Table 55: Fitting linear model: pchange ~ harmonic(pchange)

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.01159	0.002282	5.08	3.508e-06
harmonic(pchange)cos(2*pi*t)	-0.006605	0.003237	-2.041	0.04542
harmonic(pchange)sin(2*pi*t)	0.01612	0.003208	5.026	4.291e-06

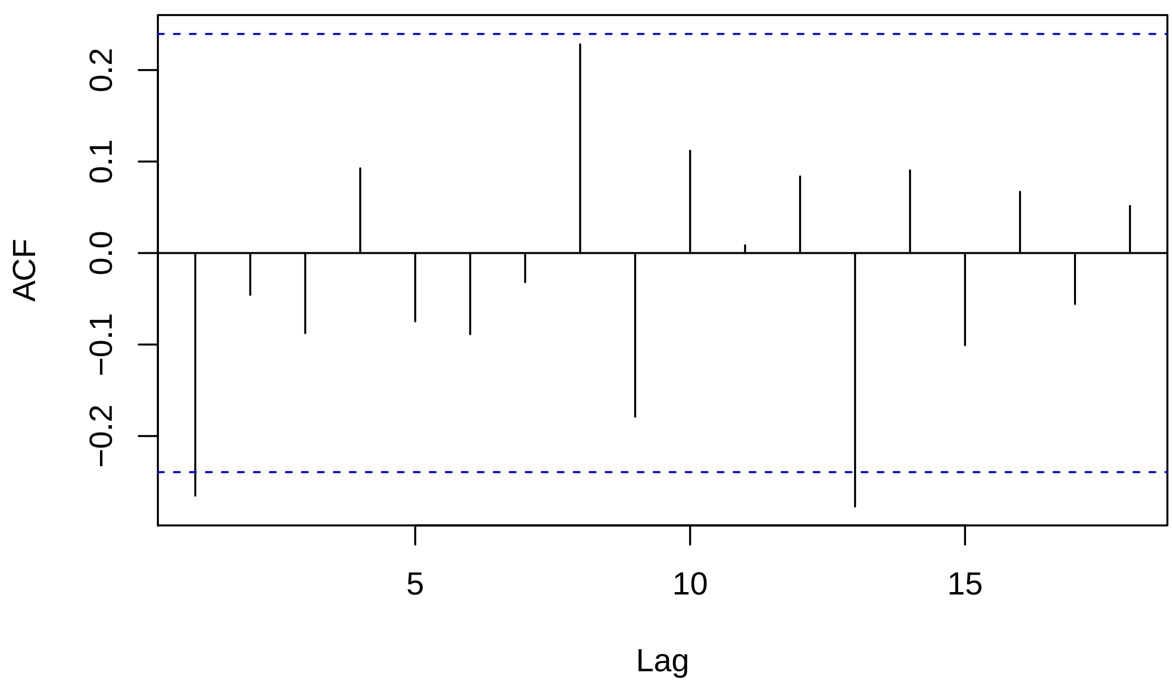
3.15b)

The runs test is significant since we have less runs than expected. We can conclude that the residuals are not random.

```
## $pvalue
## [1] 0.0026
##
## $observed.runs
## [1] 47
##
## $expected.runs
## [1] 34.43284
##
## $n1
## [1] 32
##
## $n2
## [1] 35
##
## $k
## [1] 0
```

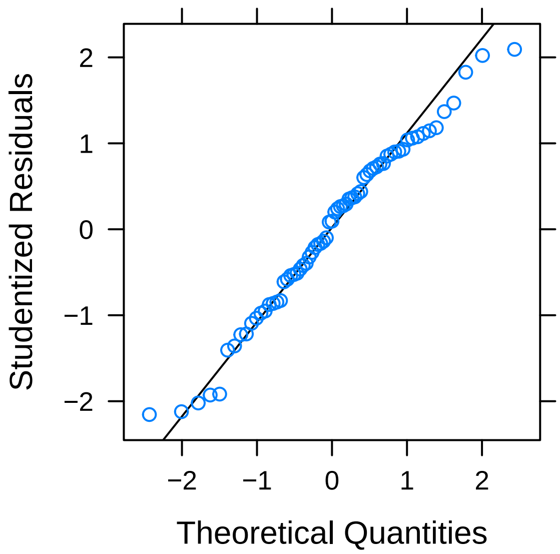
3.15c)

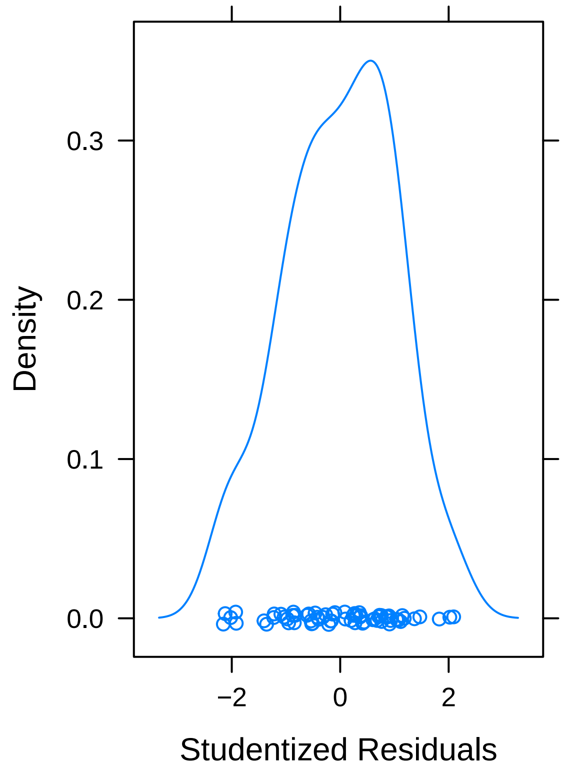
The following ACF plot indicates there are a small number of lags that have statistically significant correlation, which indicates there may be a trend that has not been taken into account. This does not appear to indicate dependency.



3.15d)

The distribution of the residuals in both the Q-Q plot and the density plot indicates a heavy-tailed and right-skewed distribution.





3.2 $Y_t = \beta_0 + \beta_1 t + X_t$

$\bar{Y} = \frac{1}{n}\sum Y_t = \frac{1}{n}\sum(\mu + e_t - e_{t-1}) = \mu + \frac{1}{n}\sum(e_t - e_{t-1}) = \mu + \frac{1}{n}(e_n - e_0)$

$Var(\bar{Y}) = Var\left(\mu + \frac{1}{n}(e_n - e_0)\right) = \frac{1}{n^2}Var(e_n - e_0) = \frac{1}{n^2}(\sigma_e^2 + \sigma_e^2) = \frac{2\sigma_e^2}{n^2}$

IT'S UNUSUAL FOR THE SAMPLE SIZE TO HAVE A BIG IMPACT ON THE VARIANCE.

$Y_t = \mu + e_t$

$\bar{Y} = \frac{1}{n}\sum Y_t = \frac{1}{n}\sum(\mu + e_t) = \mu + \frac{1}{n}\sum e_t = \mu + \bar{e}$

$Var(\bar{Y}) = Var(\mu + \bar{e}) = Var(\bar{e}) = \sigma_e^2/n$

3.3 $Y_t = \mu + e_t + e_{t-1}$

$\bar{Y} = \frac{1}{n}\sum Y_t = \frac{1}{n}\sum(\mu + e_t + e_{t-1}) = \mu + \frac{1}{n}\sum(e_t + e_{t-1})$

$= \mu + \frac{1}{n}\left(e_n + e_0 + 2\sum^{n-1} t\right)$

$Var(\bar{Y}) = Var\left(\mu + \frac{1}{n}\left(e_n + e_0 + 2\sum^{n-1} t\right)\right)$

$= \frac{1}{n^2}\left(\sigma_e^2 + \sigma_e^2 + 4(n-1)\sigma_e^2\right) = \frac{2(2n-1)\sigma_e^2}{n^2}$

$Y_t = \mu + e_t$

FROM 3.2: $Var(\bar{Y}) = \sigma_e^2/n$

AS $n \to \infty$, THE VARIANCE FROM THE ORIGINAL Y_t BECOMES 4 TIMES LARGER THAN THE SECOND Y_t