

CSSS 512 HW 1

Chris Hess

Contents

Problem 1	2
Part a	2
Part b	4
Part c	6
Part d	8
Part e	10
Part f	11
Part g	13
Part h	15
Part i	17
Part j	19
Part k	21
Part l	23
Part m	25
Part n	28
Part o	30
Part p	32
Part q	34
Part r	36
Problem 2	38
Part s	38
Part t	38
Part u	38
Part v	38
Part w	39
Problem 3	40

NB: The ACF/PACF plots start at lag = 1, not lag = 0 like with `plot(acf(ts))`

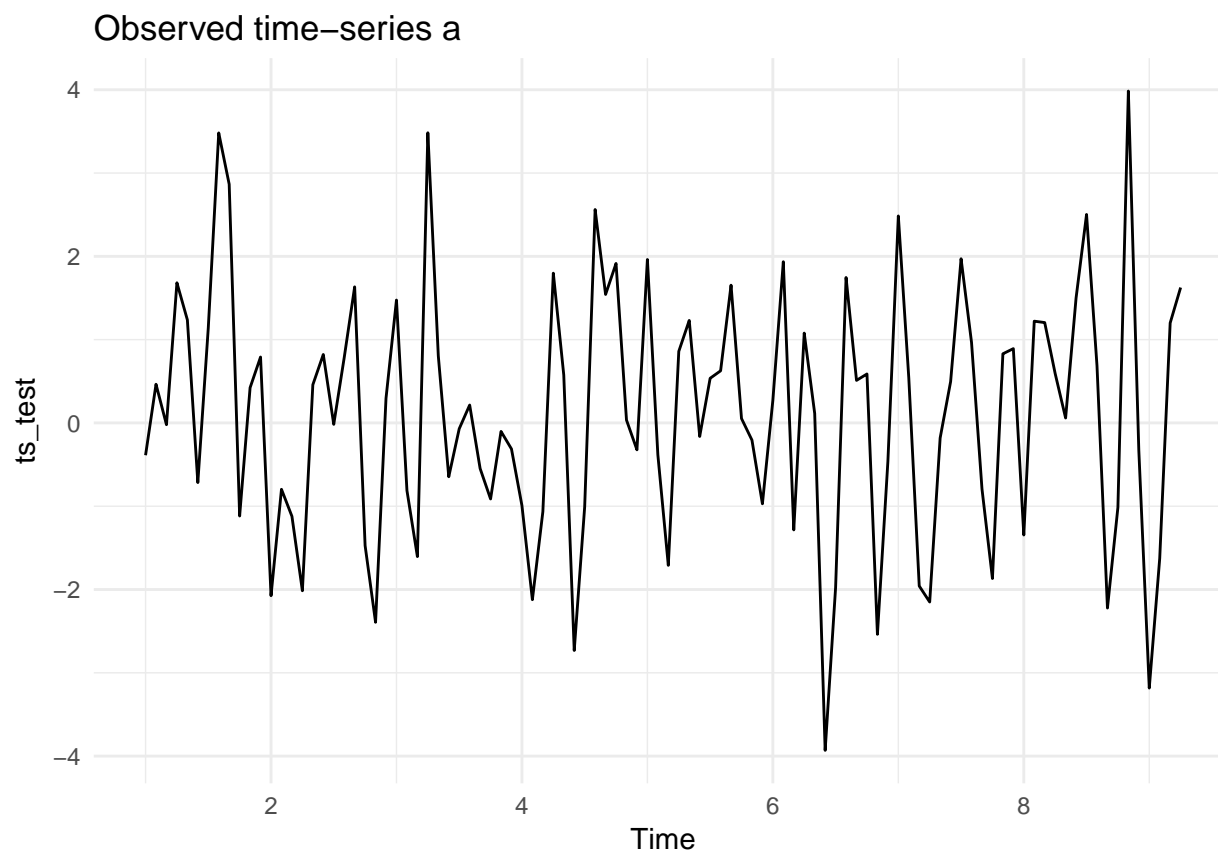
Problem 1

Part a

The time-series plot shows that the DGP reverts to its mean, but also has shocks that linger for a period or two. According to the decomposition model there's no clear seasonal pattern or trend. Looking at the ACF there's no geometric decline as would be expected in the presence of autocorrelation. An MA(2) with a small positive first-order coefficient but relatively large negative second-order coefficient captures the observed characteristic where shocks that last a period extra but do not have a lingering effect like with autoregression. The ACF and PACF of the time-series show a small positive spike for L1, but a substantial negative spike for L2 (≈ 0.5). I used `arima.sim` to simulate a MA(2), $\psi_1 = 0.15, \psi_2 = -0.5$ process and the result is similar in its temporal structure though the precise shocks and levels over time differ.

```
#pull time-series a
let <- "a"
ts_test <- ts(prob1[[let]], frequency = 12)

#time-series plot
autoplot(ts_test) +
  labs(title = "Observed time-series a") +
  theme_minimal()
```



```
#decomposition model
#autoplot(decompose(ts_test, type = "additive")) +
# theme_minimal()

#ACF
```

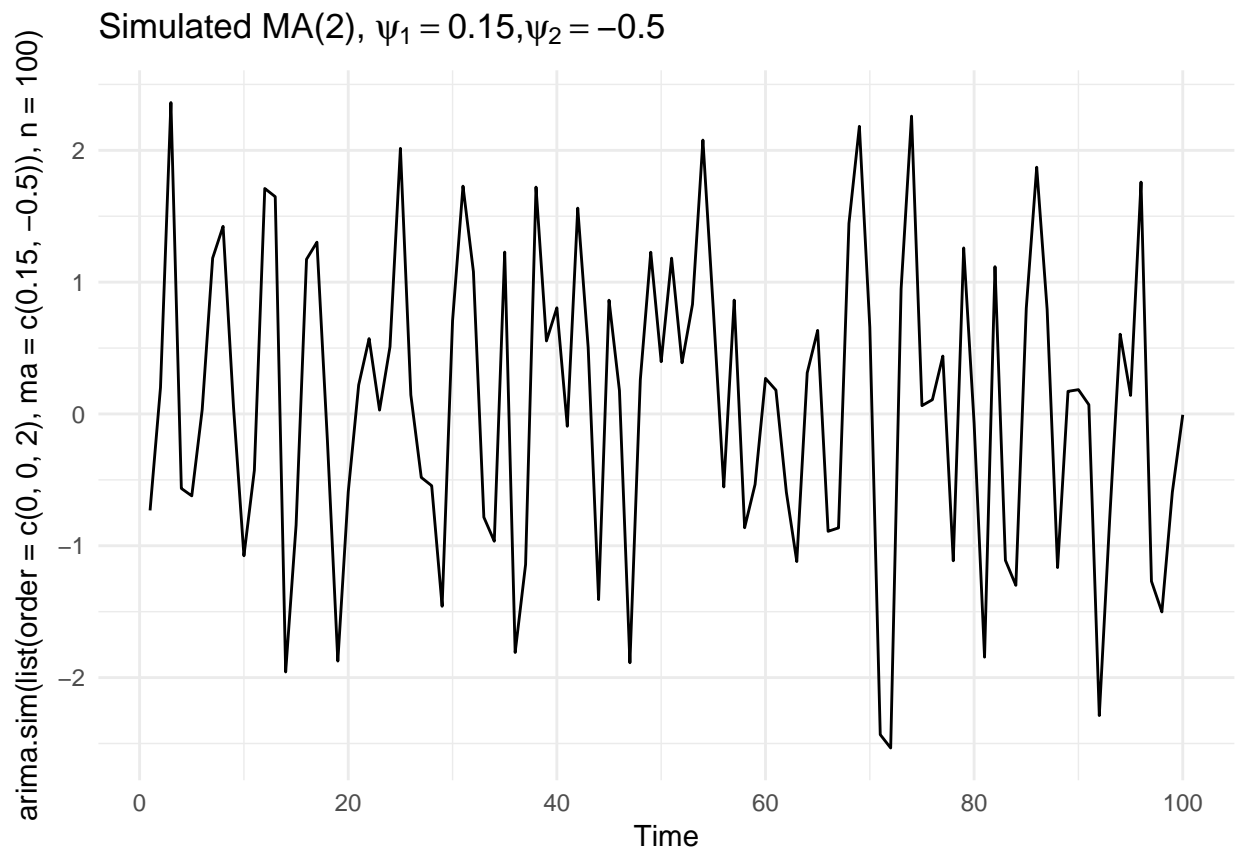
```

#ggAcf(ts_test) +
#  theme_minimal() +
#  labs(title = "")

#PACF
#ggPacf(ts_test) +
#  theme_minimal() +
#  labs(title = "")

#plot simulated ts: MA(1), psi = .75, T=100
autoplot(arima.sim(list(order = c(0, 0, 2), ma = c(0.15, -0.5)), n = 100)) +
  labs(title = TeX("Simulated MA(2),  $\psi_1 = 0.15, \psi_2 = -0.5$ ") +
  theme_minimal()

```



Part b

There appears to be a seasonal pattern looking at the time-series plot. Decomposing the time-series into its seasonal component shows that the end of each year spikes to high levels, and then reverts back to a very small/negligible seasonal effect. Plotting the ACF before removing the additive seasonal means shows a spike in autocorrelation at lag 12, corresponding to the same month of the last year. After removing the seasonal average from each observation, the ACF shows no substantial autocorrelation.

```
#pull time-series b
let <- "b"
ts_test <- ts(probi[[let]], frequency = 12)

#plot time-series
#autoplot(ts_test) +
#  theme_minimal()

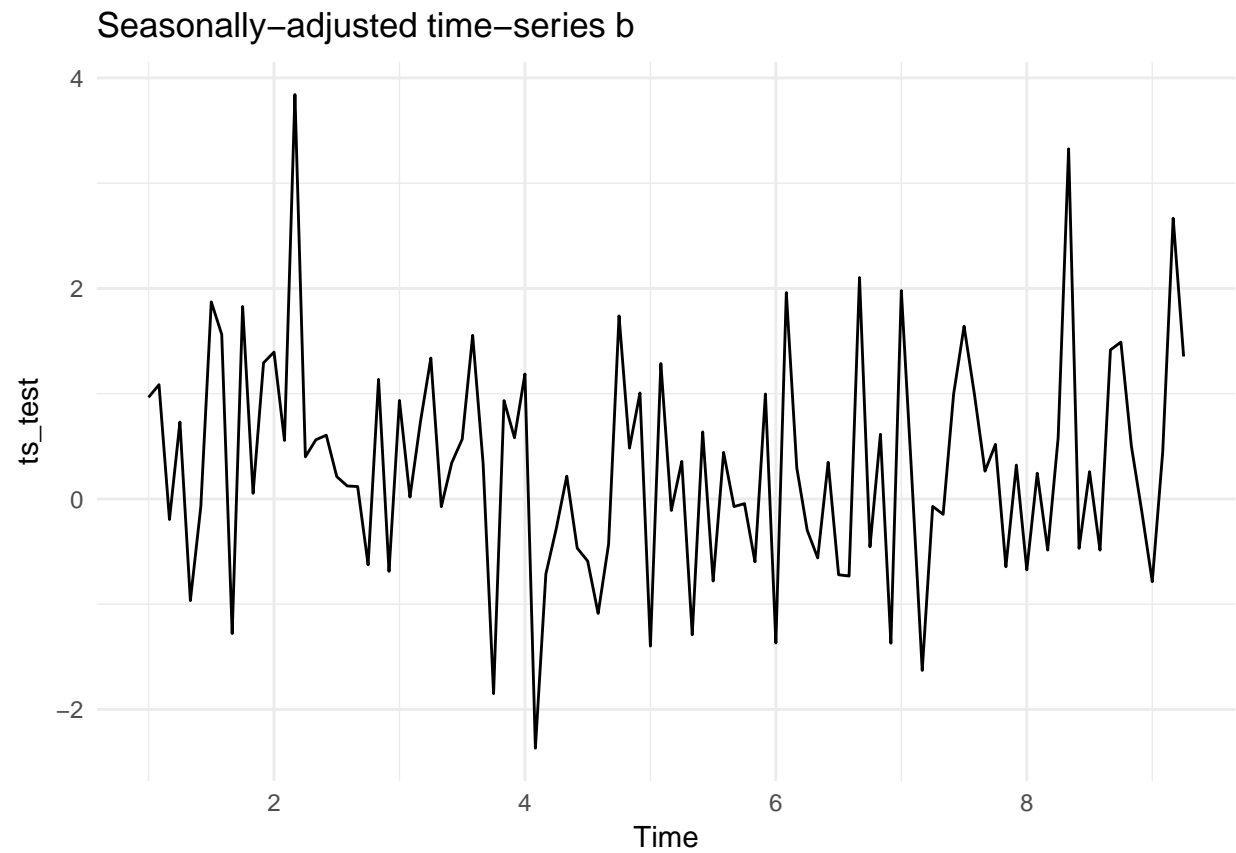
#plot decomposition model
#autoplot(decompose(ts_test, type = "additive")) +
#  theme_minimal()

#plot unadjusted ACF
#ggAcf(ts_test) +
#  theme_minimal() +
#  labs(title = "")

#store decomposition model
decomp <- decompose(ts_test, type = "additive")

#seasonally adjust (y - seasonal mean for y)
ts_test <- ts_test - decomp$seasonal

#plot adjusted time-series
autoplot(ts_test) +
  labs(title = "Seasonally-adjusted time-series b") +
  theme_minimal()
```



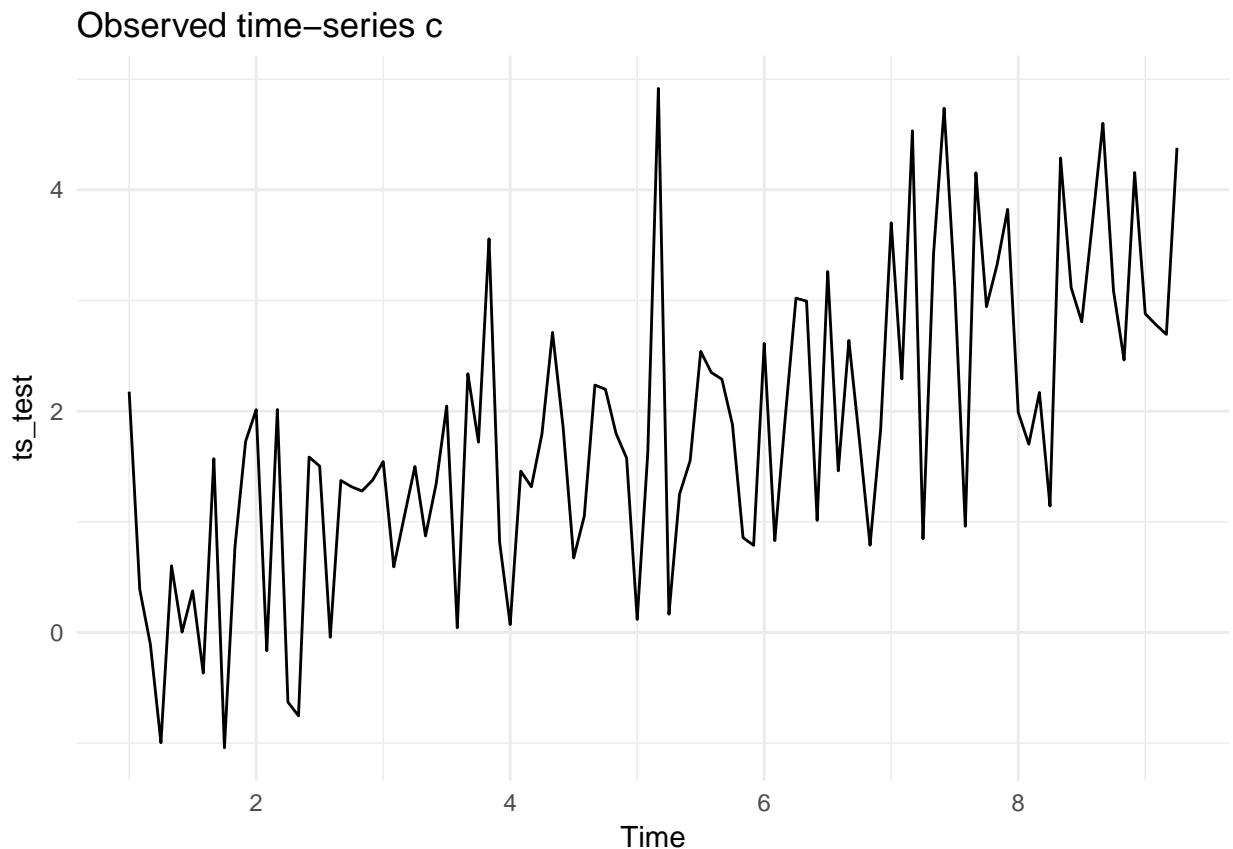
```
#plot adjusted ACF  
#ggAcf(ts_test) +  
# theme_minimal() +  
# labs(title = "")
```

Part c

There appears to be a deterministic trend according to the time-series plot and decomposition model. The ACF has some semblance to an AR process, albeit without a clearly geometric decline in autocorrelation as might be expected. Detrending the data using a linear model shows that the removing deterministic trend explains away the early trappings of autoregression (i.e. the detrended time-series looks like white noise, ACF has no structure).

```
#pull time-series c
let <- "c"
ts_test <- ts(probl[[let]], frequency = 12)

#plot time-series
autoplot(ts_test) +
  labs(title = "Observed time-series c") +
  theme_minimal()
```



```
#plot decomposition model
#autoplot(decompose(ts_test, type = "additive")) +
#  theme_minimal()

#plot observed ACF
#ggAcf(ts_test) +
#  theme_minimal() +
#  labs(title = "")

#looks like trend, let's estimate the monthly trend
```

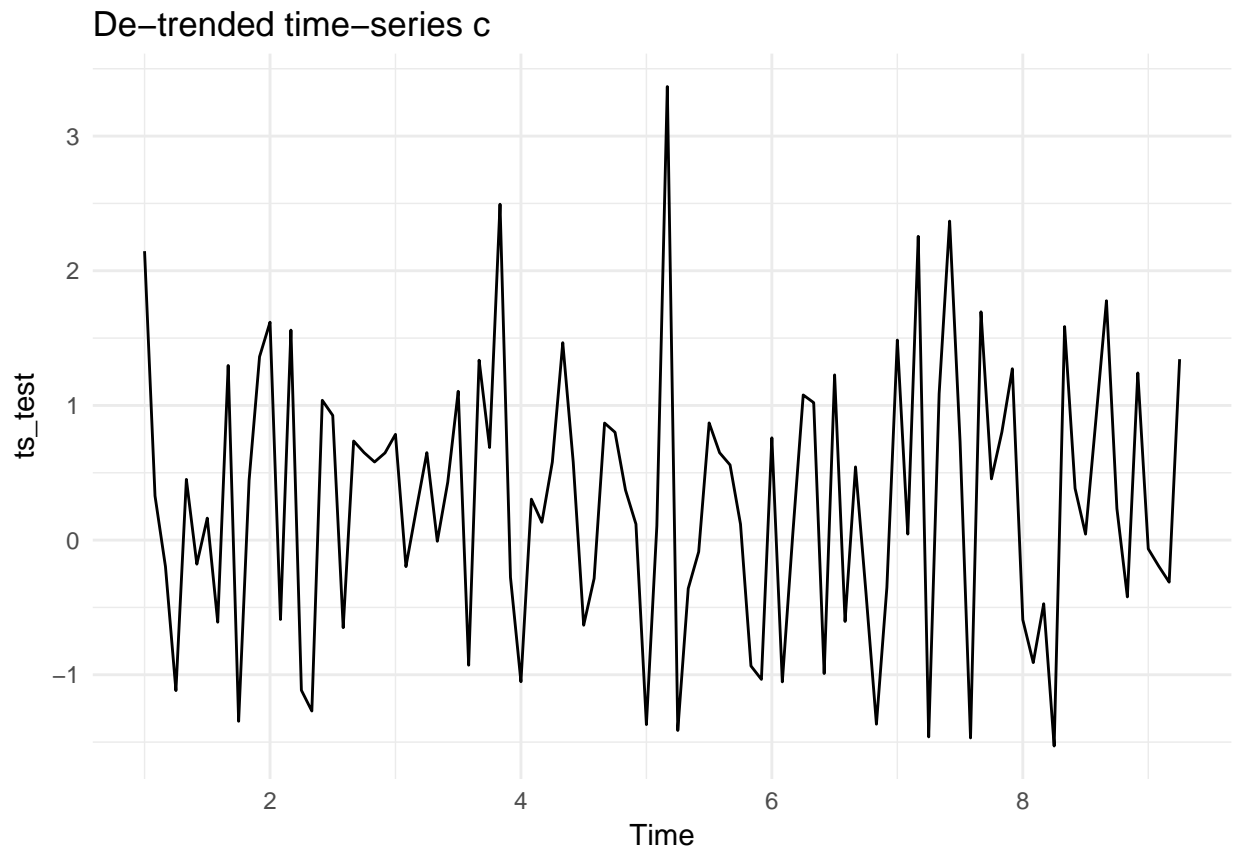
```

time <- 1:100
lm_c <- lm(ts_test ~ time)
#summary(lm_c) #about .03 increase per month

#subtract estimated trend from time-series
ts_test <- ts_test - lm_c$coefficients[2]*time

#plot de-trended series
autoplot(ts_test) +
  labs(title = "De-trended time-series c") +
  theme_minimal()

```



```

#plot de-trended ACF
#ggAcf(ts_test) +
#  labs(title = "De-trended ACF for time-series c") +
#  theme_minimal() +
#  labs(title = "")

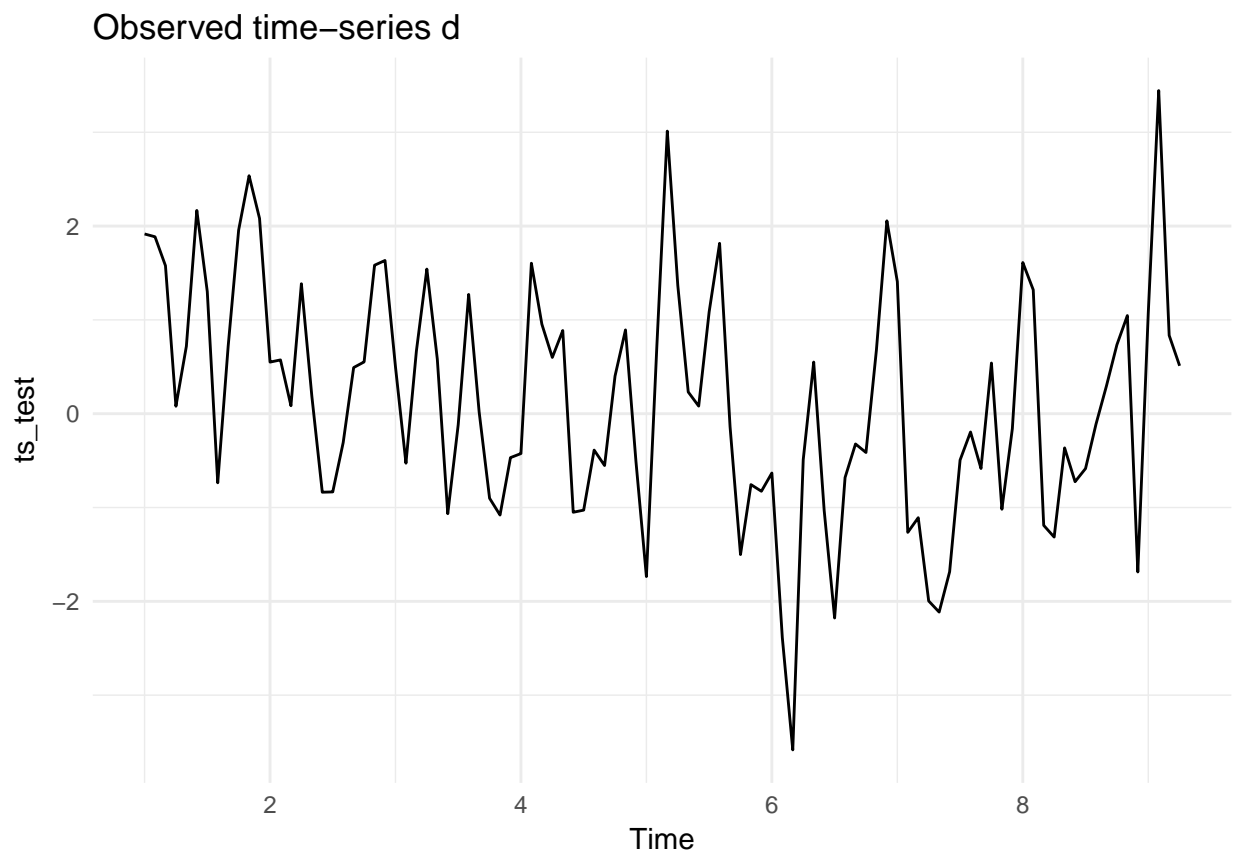
```

Part d

There seems to be an autoregressive or moving-average process using the time-series plot as a first-look, but there is no trend. The decomposition model also suggests no trend, and though the seasonal pattern is plausible (i.e. summer months dip, rises through fall) in the absence of any notion of what the underlying time-series it does not seem to explain much overall variation in the series. While plotting the ACF of the time-series shows a spike at L1, the effect of the prior period dies off quickly (evidence in favor of MA). The PACF shows the L1 spike is about 0.5, though there are also spikes at L2 and L3 (with negative and positive signs, respectively). Based off of the inconsistent evidence of autocorrelation and quick mean reversion, a moving average process seems most likely. I simulated an MA(1) with $\psi = 0.5$, MA(2) with $\psi_1 = 0.5, \psi_2 = -0.25$ to investigate the DGP further. Both have a positive first-order term because a shock in the prior period lingers for another period, but the MA(2) has a second-order negative term driving the time-series back to the mean faster. Based off of my inspection of these simulations I think the MA(1) with $\psi = 0.5$ is most appropriate.

```
#pull time-series d
let <- "d"
ts_test <- ts(prob1[[let]], frequency = 12)

#plot time-series
autoplot(ts_test) +
  theme_minimal() +
  labs(title = "Observed time-series d")
```



```
#plot decomposition model
#autoplot(decompose(ts_test, type = "additive")) +
# theme_minimal()
```



```

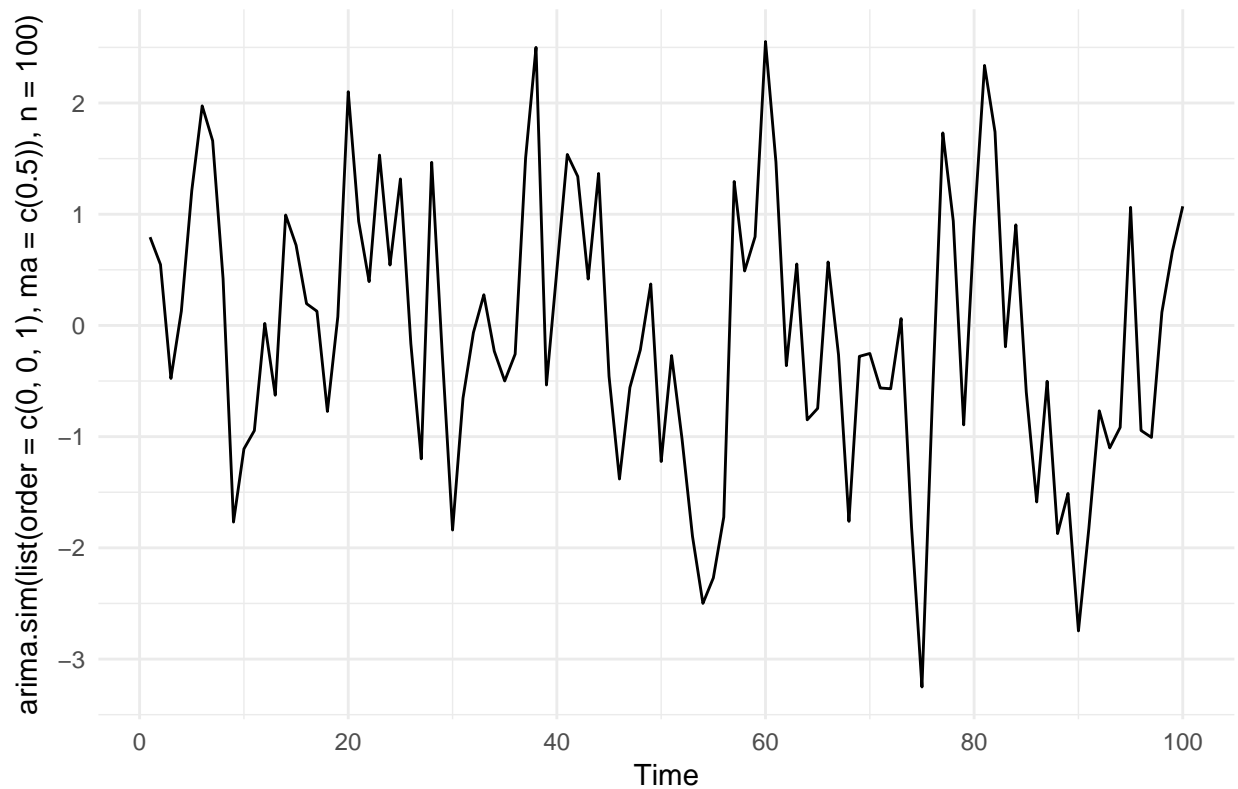
#plot ACF
#ggAcf(ts_test) +
# theme_minimal() +
# labs(title = "")

#plot PACF
#ggPacf(ts_test) +
# theme_minimal() +
# labs(title = "")

#plot simulated ts: MA(1), psi = .5, T=100
autoplot(arima.sim(list(order = c(0, 0, 1), ma = c(0.5)), n = 100)) +
  theme_minimal() +
  labs(title = TeX("Simulated MA(1),  $\psi_1 = 0.5$ "))

```

Simulated MA(1), $\psi_1 = 0.5$



```

#plot simulated ts: MA(2), phi = .5, T=100
#autoplot(arima.sim(list(order = c(0, 0, 2), ma = c(0.5, -0.25)), n = 100)) +
# theme_minimal()

```

Part e

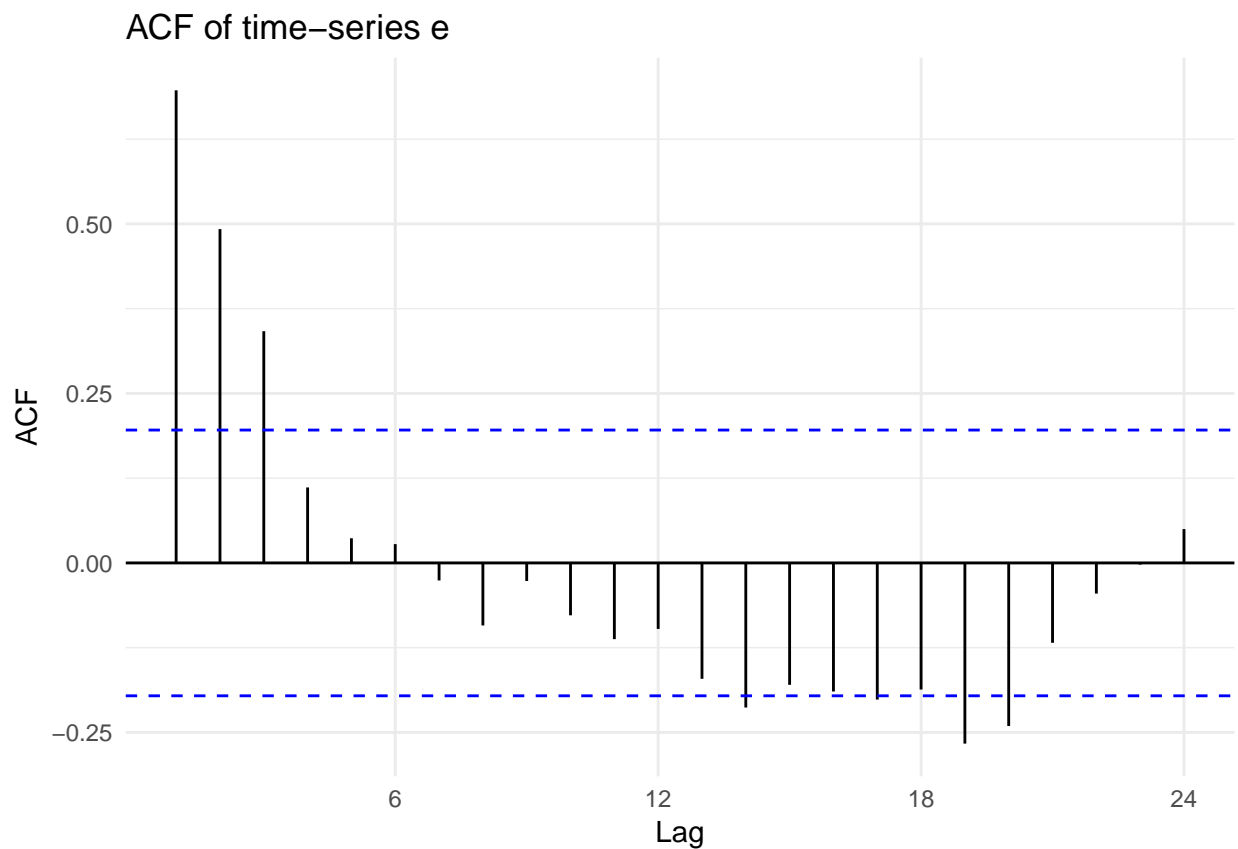
The time-series plot gives pretty clear indication of autocorrelation, since high values beget high values, while low values similarly follow preceding low values. The decomposition model indicates an unlikely U-shaped seasonal variation that otherwise does not fit the time-series, also shows no evidence of trend. Plotting the ACF, however, shows a geometric decline in autocorrelation that extends out to about L3 at substantial levels. The PACF suggests that one lag is sufficient for the series' autocorrelation, and that the AR(1) has a plausible coefficient of $\phi = 0.7$

```
#pull time-series e
let <- "e"
ts_test <- ts(prob1[[let]], frequency = 12)

#plot time-series e
#autoplot(ts_test) +
# theme_minimal()

#autoplot(decompose(ts_test, type = "additive")) +
# theme_minimal()

ggAcf(ts_test) +
  theme_minimal() +
  labs(title = "ACF of time-series e")
```



```
#ggPacf(ts_test) +
# theme_minimal() +
# labs(title = "")
```

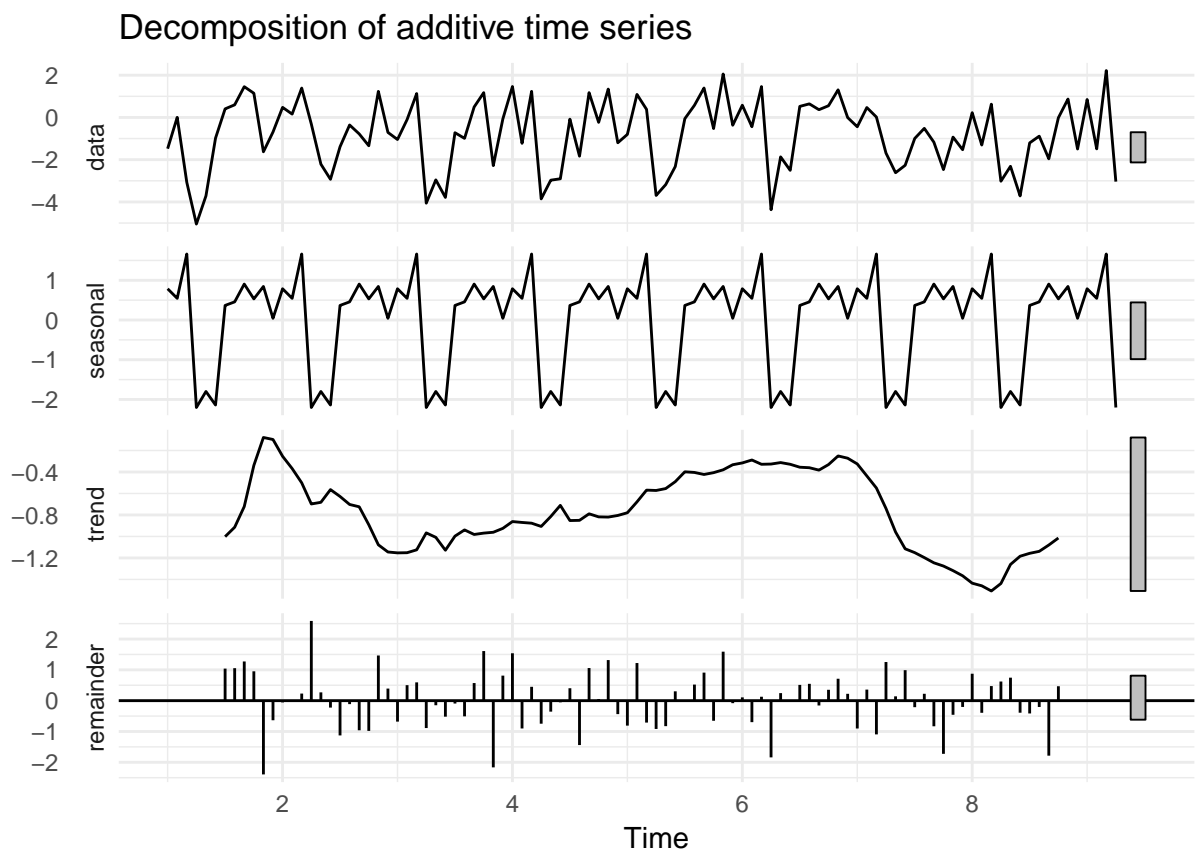
Part f

The time-series plot does not indicate much by itself, though combined with the decomposition model it is a little clearer that there is some amount of seasonality to each year observed in the time-series. After a bump upward in the first three months there is a sharp dip downward for three monthly periods followed by a return to relatively normal levels for the next 6 months. After adjusting the time-series with the monthly averages, the data look mostly like white noise. The ACF for the adjusted series show a modest negative spike in autocorrelation at L12, but it is reasonable that this could be due to chance too (and this is suspected given the provided information about the time-series).

```
#pull time-series f
let <- "f"
ts_test <- ts(prob1[[let]], frequency = 12)

#plot time-series
#autoplot(ts_test) +
#  theme_minimal()

#plot decomposition model
autoplot(decompose(ts_test, type = "additive")) +
  theme_minimal()
```



```
#plot ACF
#ggAcf(ts_test) +
#  theme_minimal() +
#  labs(title = "")
```

```

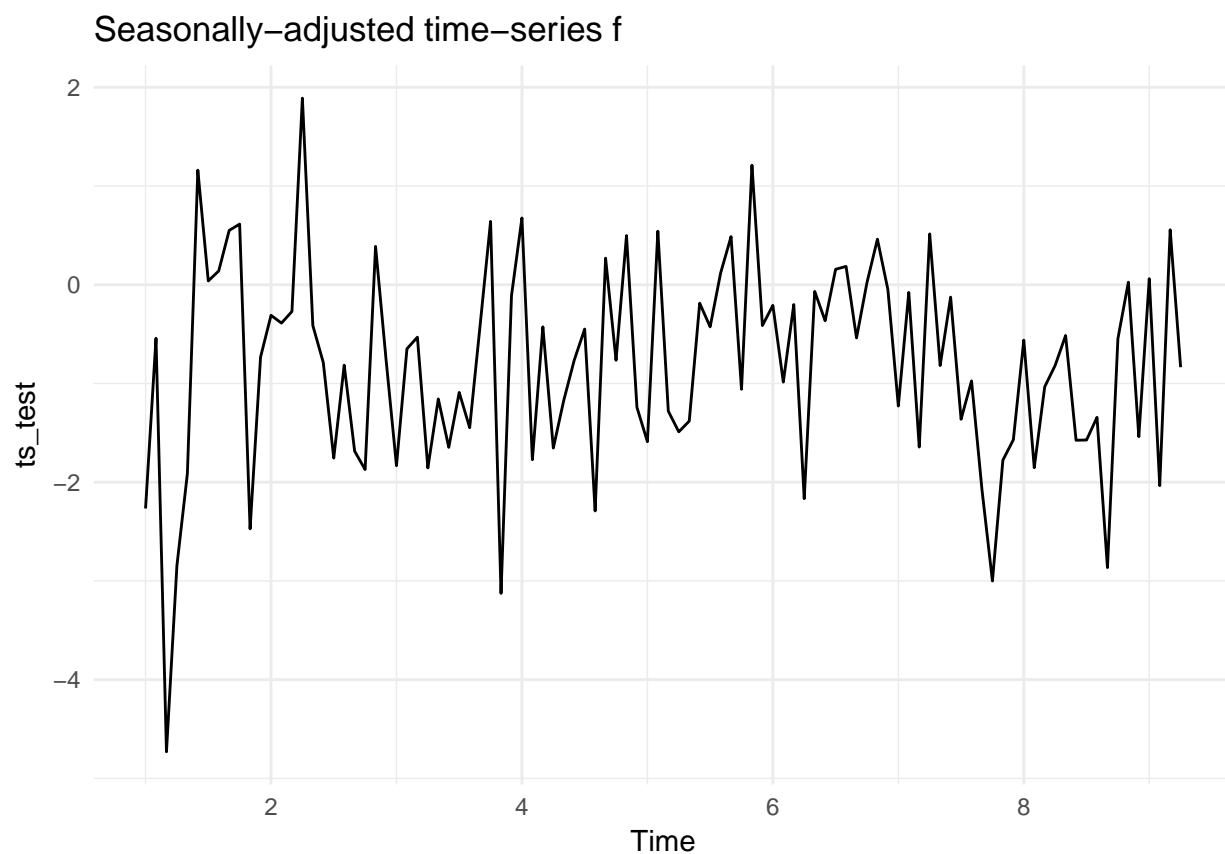
#plot PACF
#ggPacf(ts_test) +
# theme_minimal() +
# labs(title = "")

#store decomposition model
decomp <- decompose(ts_test, type = "additive")

#seasonally adjust (y - seasonal mean for y)
ts_test <- ts_test - decomp$seasonal

#plot time-series
autoplot(ts_test) +
  labs(title = "Seasonally-adjusted time-series f") +
  theme_minimal()

```



```

#ggAcf(ts_test) +
# labs(title = "Seasonally-adjusted ACF for time-series f") +
# theme_minimal()

```

Part g

This time-series seems to be an AR(1) process, particularly given the ACF and PACF plots. The plotted time-series shows that the time-series will stay outward from the mean for a while (positive in the middle of the series, negative just after). The ACF shows a geometric decline in autocorrelation as would be expected in the presence of an AR(p) process. Further, the PACF only has a substantial spike at L1 (the very high order ones with values above $|0.25|$ are inconsistent) and reaches a PACF value of about .75. I simulated a AR(1) with $\phi = 0.75$ and given the PACF value for L1, and the resulting time series has similar characteristics.

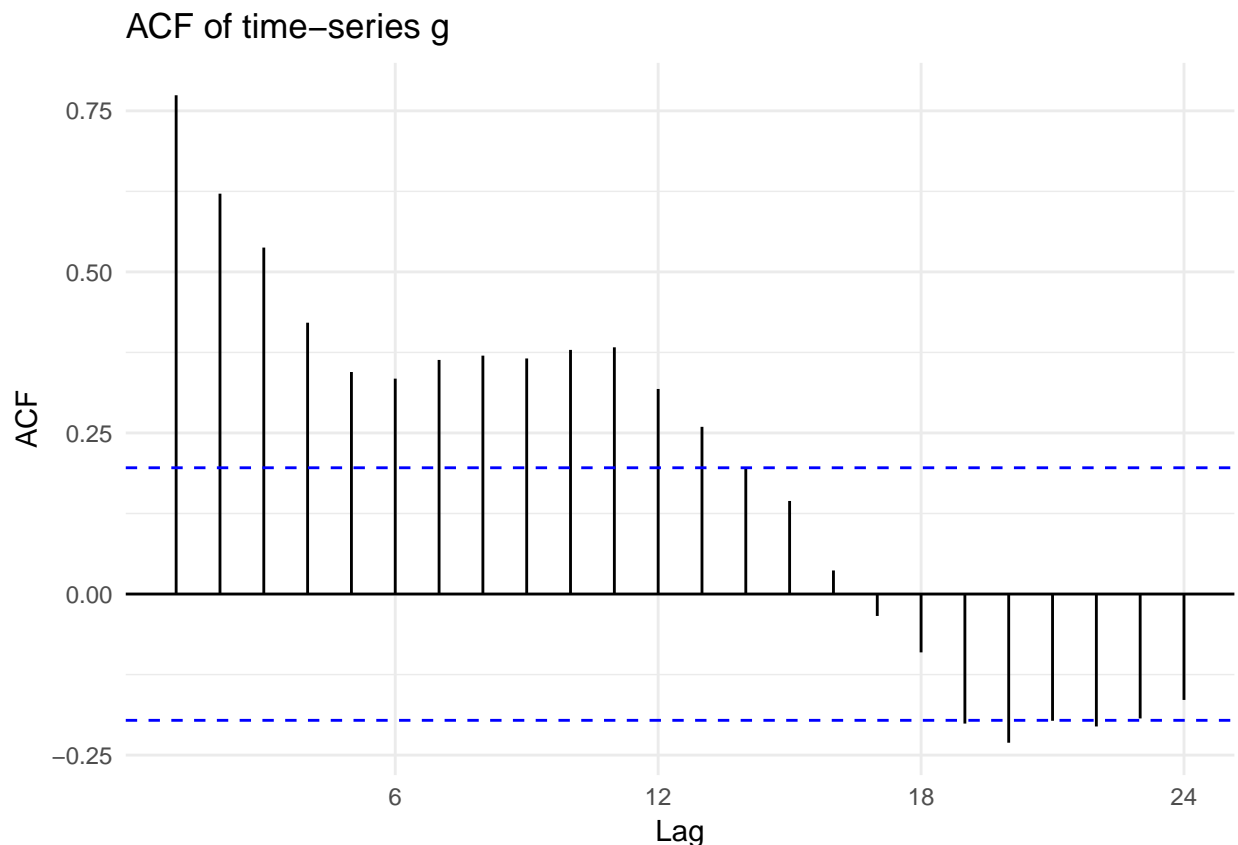
```
let <- "g"

ts_test <- ts(prob1[[let]], frequency = 12)

#autoplot(ts_test) +
#  theme_minimal()

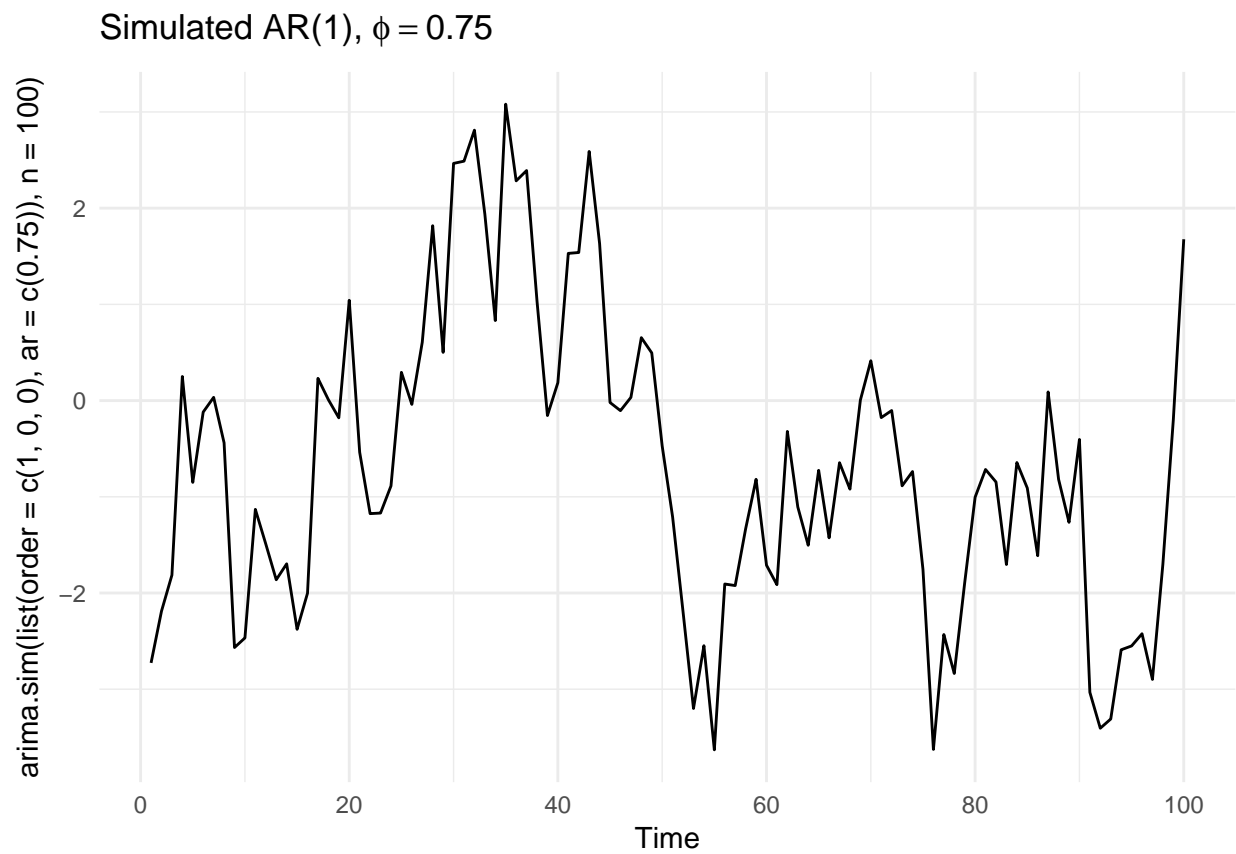
#autoplot(decompose(ts_test, type = "additive")) +
#  theme_minimal()

ggAcf(ts_test) +
  theme_minimal() +
  labs(title = "ACF of time-series g")
```



```
#ggPacf(ts_test) +
#  theme_minimal() +
#  labs(title = "")
```

```
autoplot(arima.sim(list(order = c(1, 0, 0), ar = c(0.75)), n = 100)) +  
  theme_minimal() +  
  labs(title = TeX("Simulated AR(1),  $\phi = 0.75$ "))
```

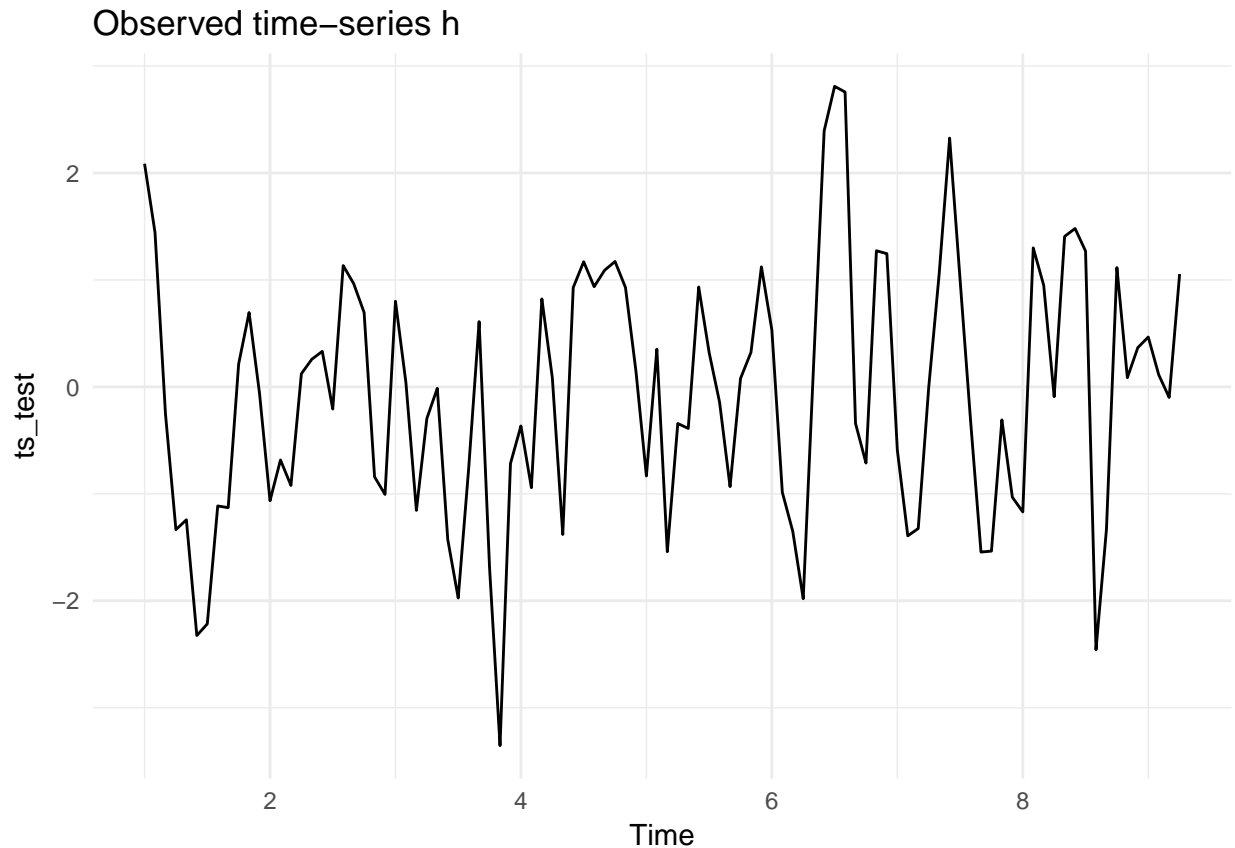


Part h

The plotted time-series does not indicate a trend, seasonal structure or much notion of autocorrelation since shocks die out fairly quickly. This leads to a series that mostly stays around its means, though the magnitude of the shocks can be fairly strong. The ACF only has a significant lag at L1, so the cumulative evidence points to MA as the temporal characteristic of the DGP. I think a MA(2) with a moderate positive first-order (0.4) and weaker negative second-order (-0.2) captures the observed time-series h because there are pretty large shocks away from the mean, though they revert back within a few periods.

```
#pull time-series h
let <- "h"
ts_test <- ts(prob1[[let]], frequency = 12)

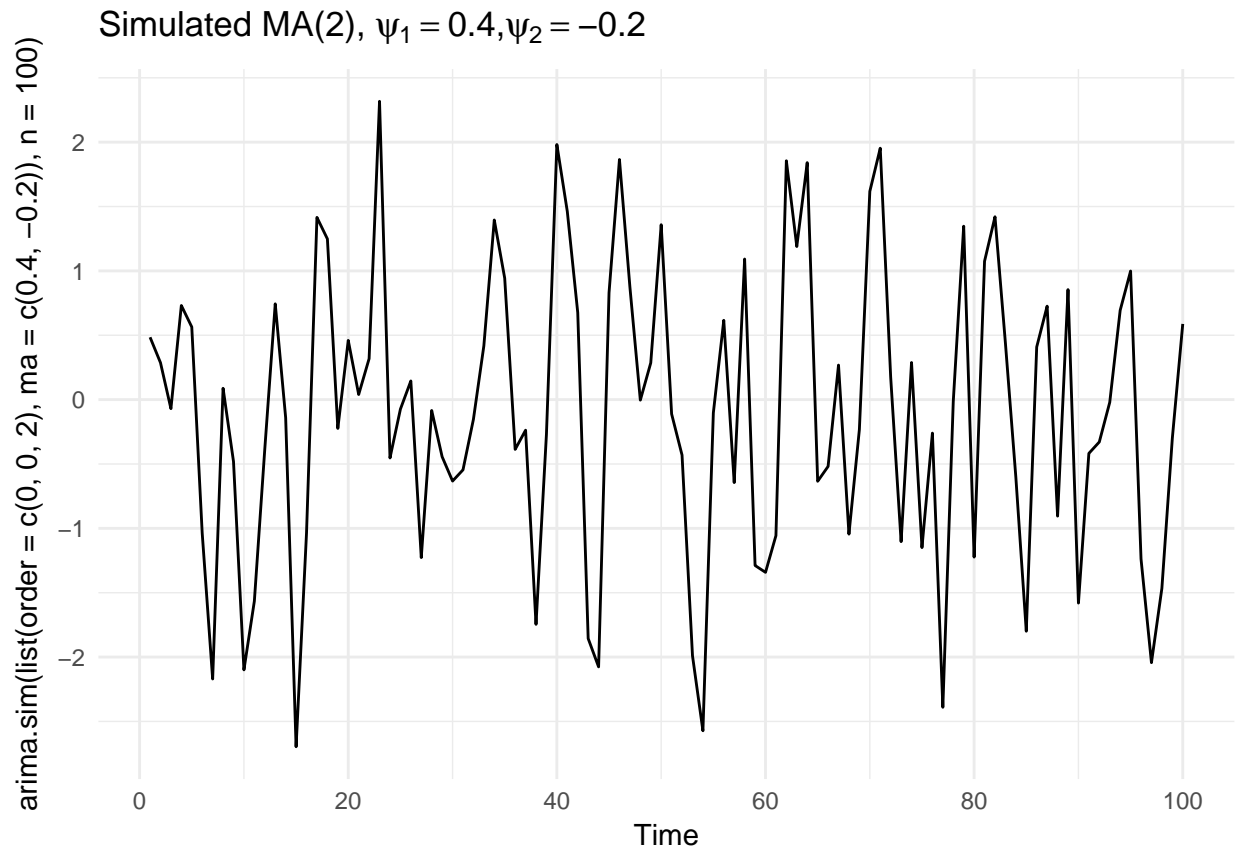
#plot the time-series
autoplot(ts_test) +
  theme_minimal() +
  labs(title = "Observed time-series h")
```



```
#plot the decomposition
#autoplot(decompose(ts_test, type = "additive")) +
#  theme_minimal()

#plot the ACF
#ggAcf(ts_test) +
#  theme_minimal() +
#  labs(title = "")
```

```
#plot simulated MA(2)
autoplot(arima.sim(list(order = c(0, 0, 2), ma = c(0.4, -0.2)), n = 100)) +
  theme_minimal() +
  labs(title = TeX("Simulated MA(2),  $\psi_1 = 0.4, \psi_2 = -0.2$ "))
```

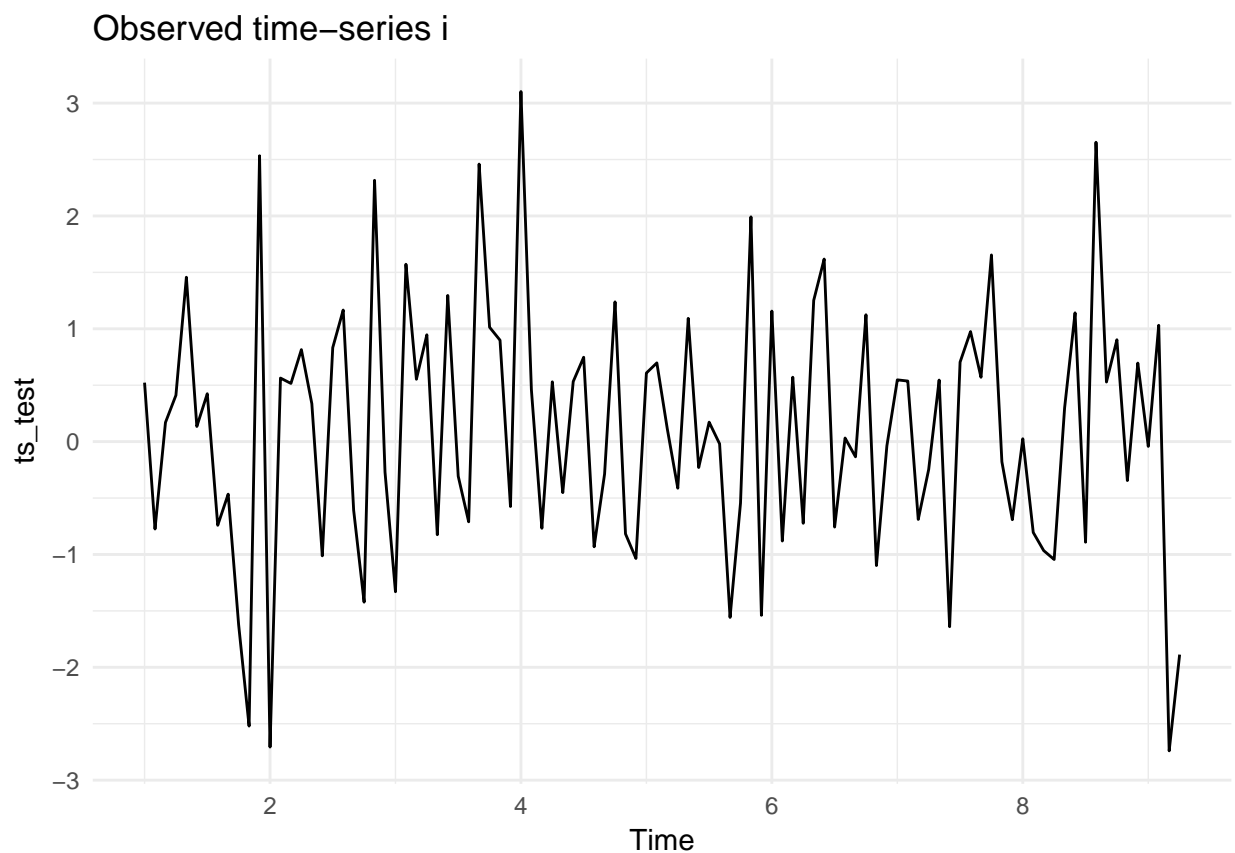


Part i

This time-series does not scream one type of process over the other as far as AR or MA are concerned, though there is clearly no trend or cycle. The ACF plot suggests a weak negative AR(1) process (-0.2) would be most appropriate. The PACF has a significant lag at L6, but this seems likely to be due to chance since there's not much of a cyclical structure otherwise.

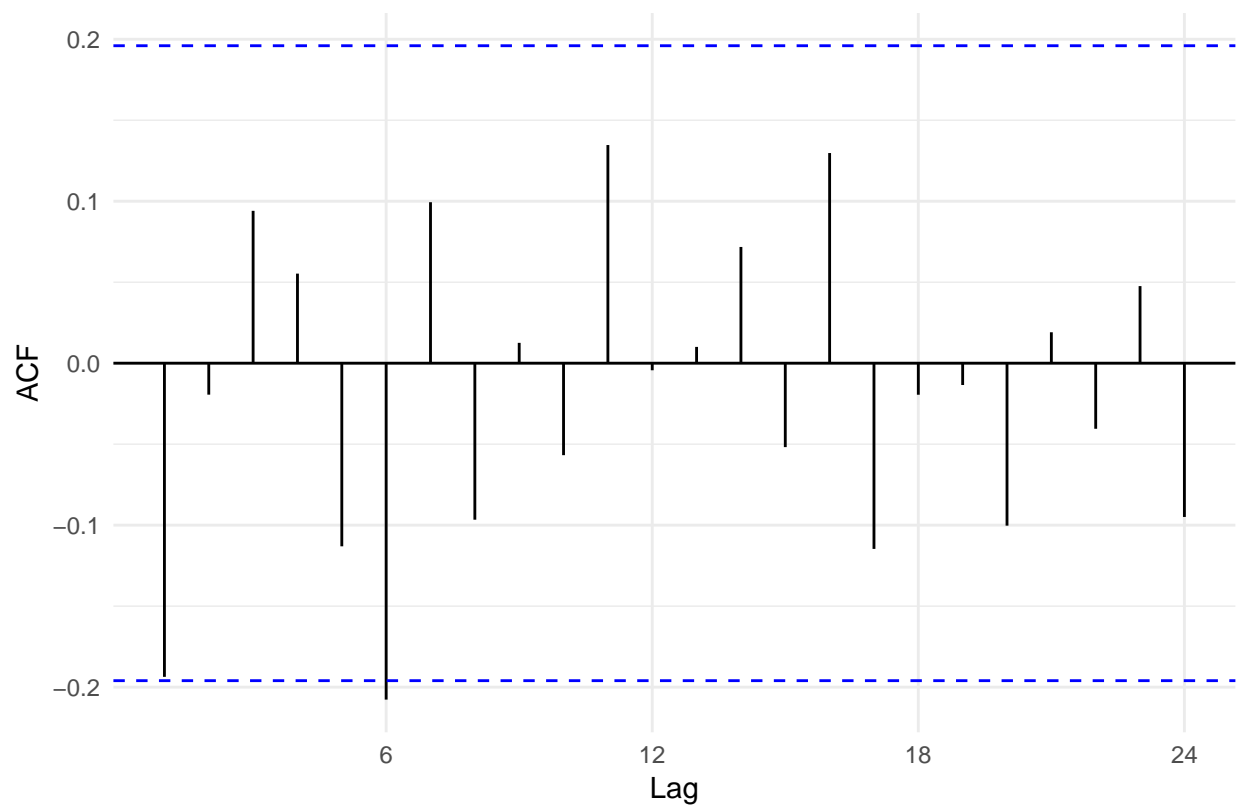
```
#pull time-series i
let <- "i"
ts_test <- ts(probl[[let]], frequency = 12)

#plot the time-series
autoplot(ts_test) +
  theme_minimal() +
  labs(title = "Observed time-series i")
```



```
#plot the decomposition
#autoplot(decompose(ts_test, type = "additive")) +
#  theme_minimal()

#plot the ACF
ggAcf(ts_test) +
  theme_minimal() +
  labs(title = "")
```



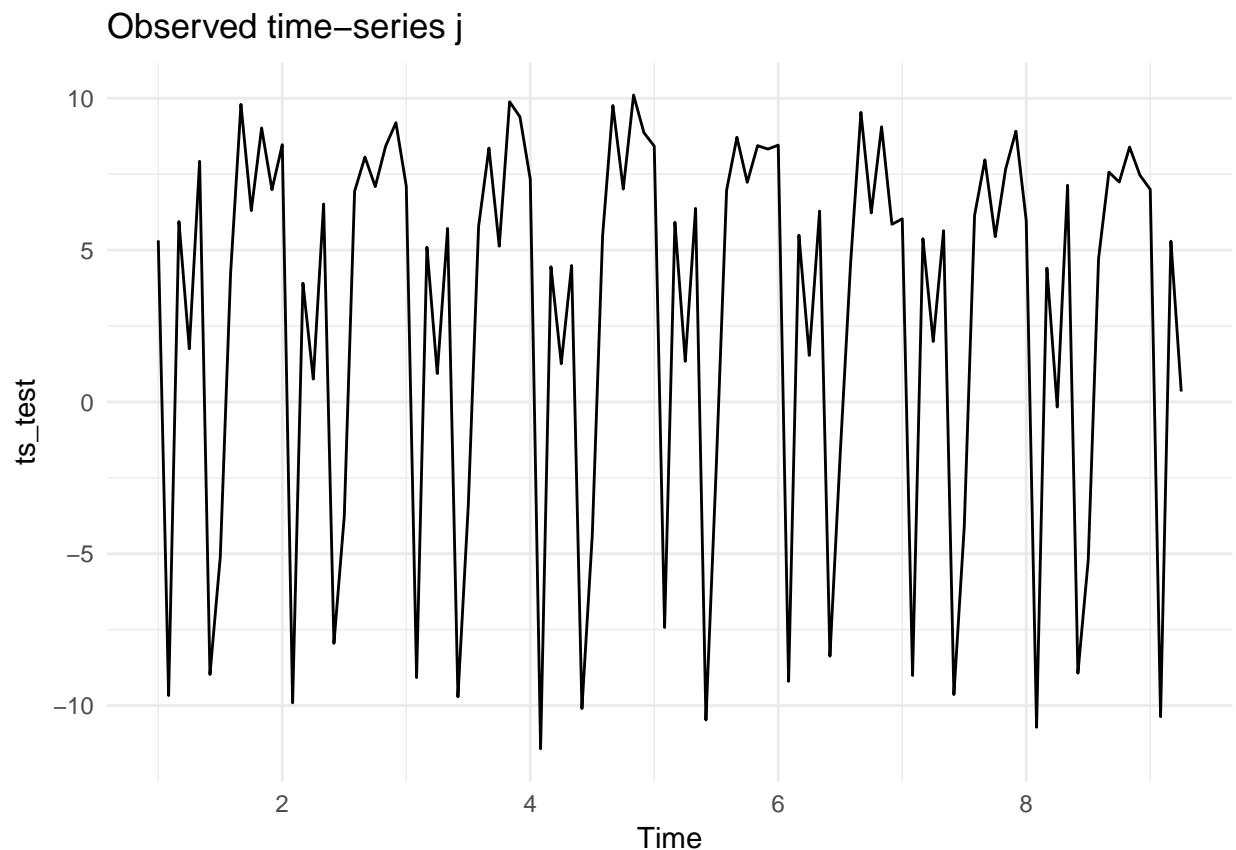
```
#plot the PACF  
#ggPacf(ts_test) +  
# theme_minimal() +  
# labs(title = "")
```

Part j

Plotting time-series j shows a clear seasonal structure with no trend. The decomposition model shows a dip in the second month of each year to relatively low levels, then the next three months have higher levels, then there is a nother slightly longer dip and lastly a return to normal levels again. Though this seasonal structure is a little odd, it doesn't seem implausible if there are a couple months of the year where levels are just exceptionally low. The strong spike in autocorrelation at L12 in the ACF plot also corroborates the seasonal structure. After adjusting with the estimated seasonal means, the time-series mostly looks like white noise.

```
#pull time-series j
let <- "j"
ts_test <- ts(probl[[let]], frequency = 12)

#plot the time-series
autoplot(ts_test) +
  labs(title = "Observed time-series j") +
  theme_minimal()
```



```
#plot the decomposition
#autoplot(decompose(ts_test, type = "additive")) +
#  theme_minimal()

#plot the ACF
#ggAcf(ts_test) +
#  theme_minimal() +
#  labs(title = "")
```

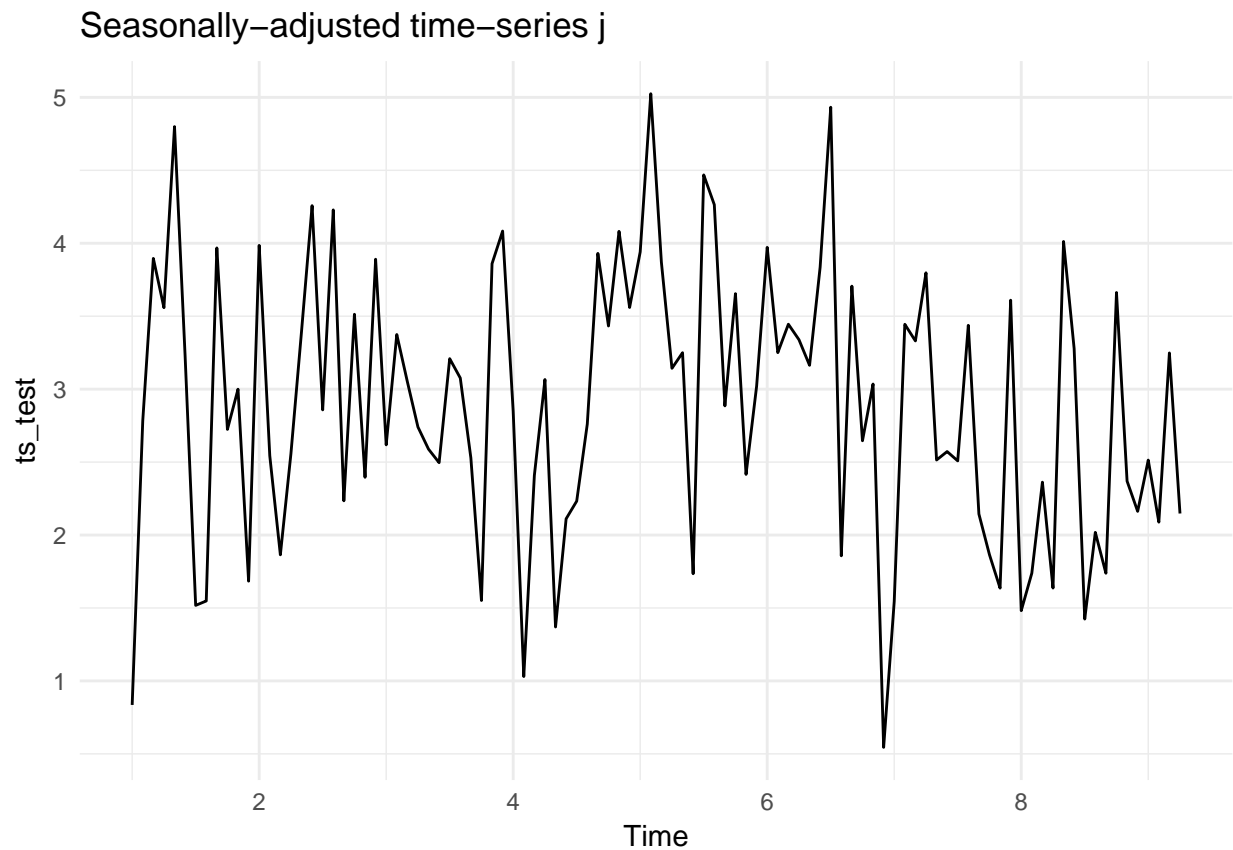
```

#store decomposition model
decomp <- decompose(ts_test, type = "additive")

#seasonally adjust (y - seasonal mean for y)
ts_test <- ts_test - decomp$seasonal

#plot adjusted time-series
autoplot(ts_test) +
  labs(title = "Seasonally-adjusted time-series j") +
  theme_minimal()

```



```

#plot the adjusted ACF
#ggAcf(ts_test) +
#  theme_minimal() +
#  labs(title = "")

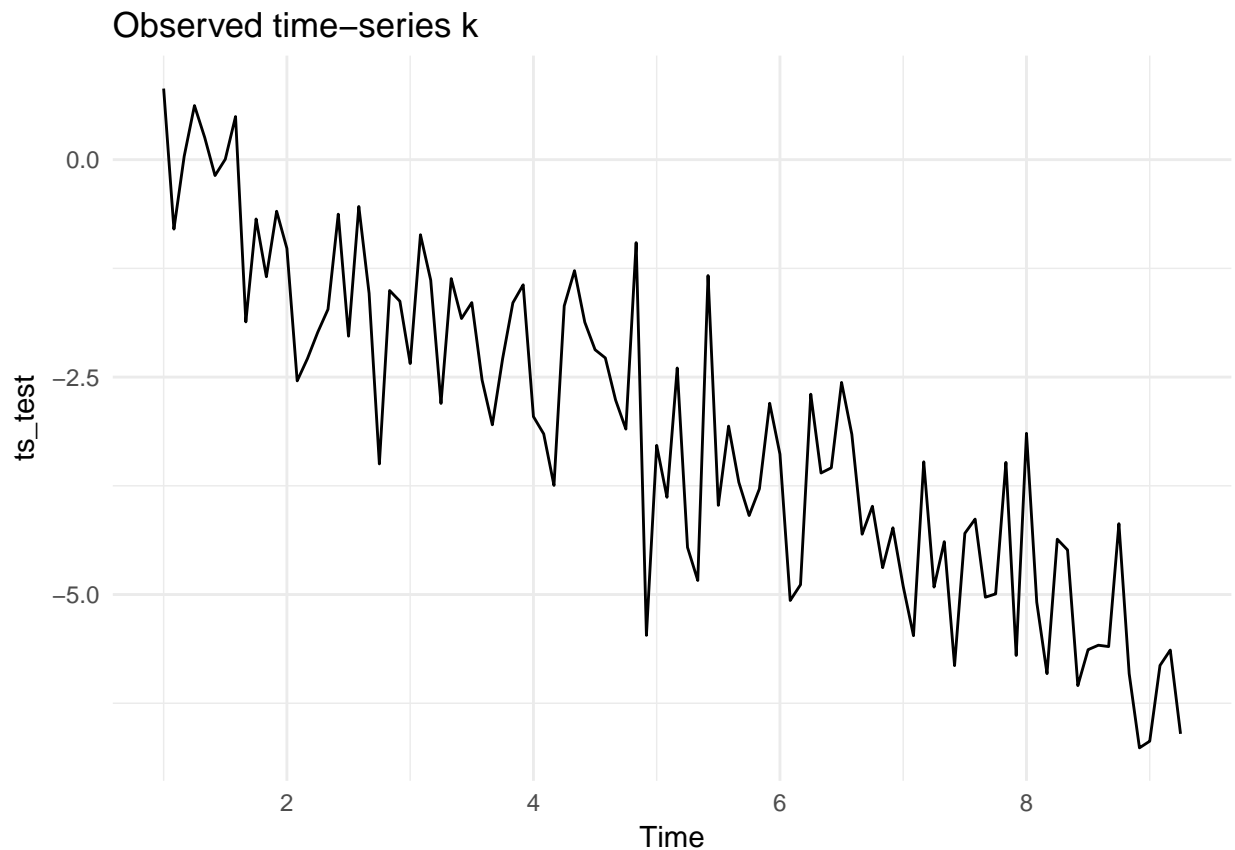
```

Part k

The time-series plot clearly shows a deterministic negative trend, which the decomposition model also points to. After detrending the time-series, the resulting series is mostly white noise. There was early evidence of autocorrelation in the unadjusted series, but after detrending the ACF shows no substantial autocorrelation at any lag.

```
#pull time series k
let <- "k"
ts_test <- ts(probi[[let]], frequency = 12)

#plot the time-series
autoplot(ts_test) +
  labs(title = "Observed time-series k") +
  theme_minimal()
```



```
#plot the decomposition
#autoplot(decompose(ts_test, type = "additive")) +
# theme_minimal()

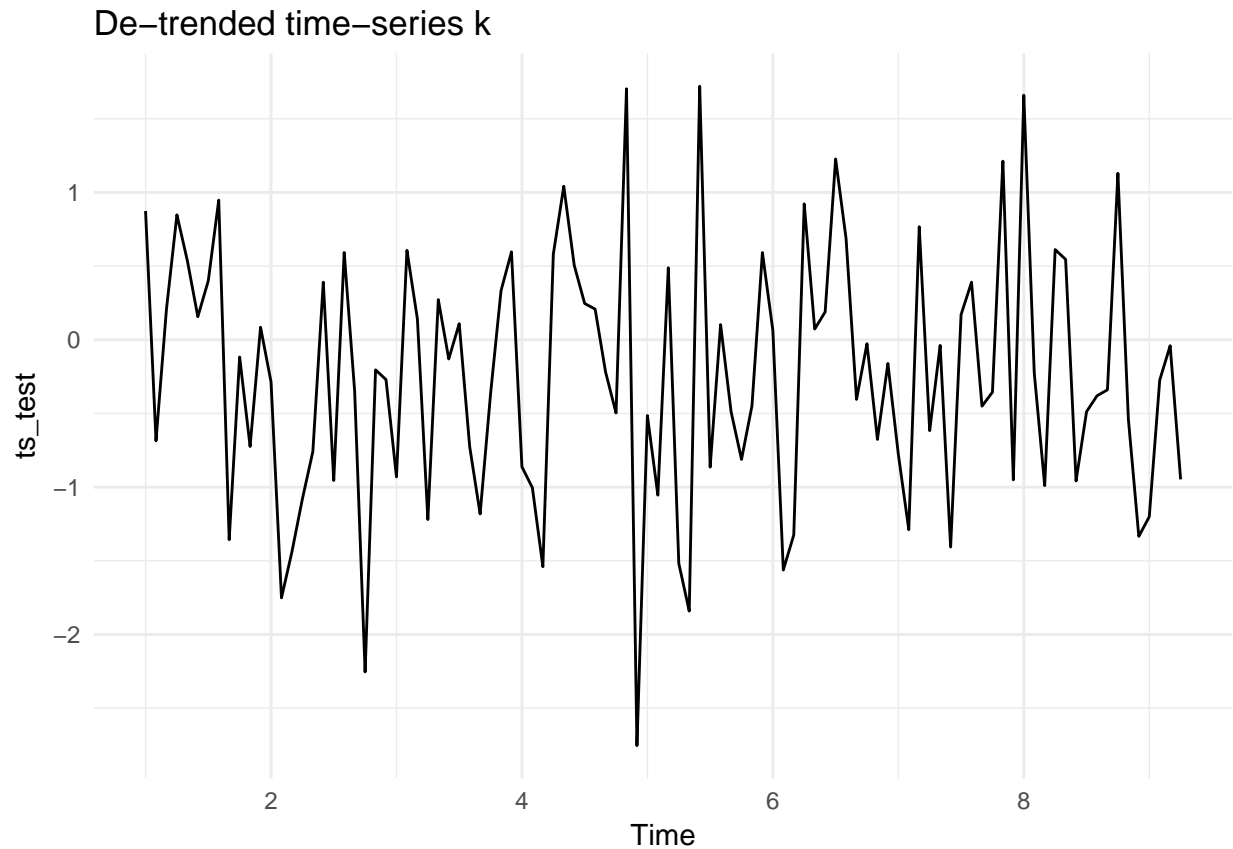
#plot the ACF
#ggAcf(ts_test) +
# theme_minimal() +
# labs(title = "")

#looks like trend, let's estimate the monthly trend
time <- 1:100
```

```
lm_k <- lm(ts_test ~ time)
#summary(lm_k) #about -.05 increase per month

#subtract estimated trend from time-series
ts_test <- ts_test - lm_k$coefficients[2]*time

#plot de-trended series
autoplot(ts_test) +
  labs(title = "De-trended time-series k") +
  theme_minimal()
```



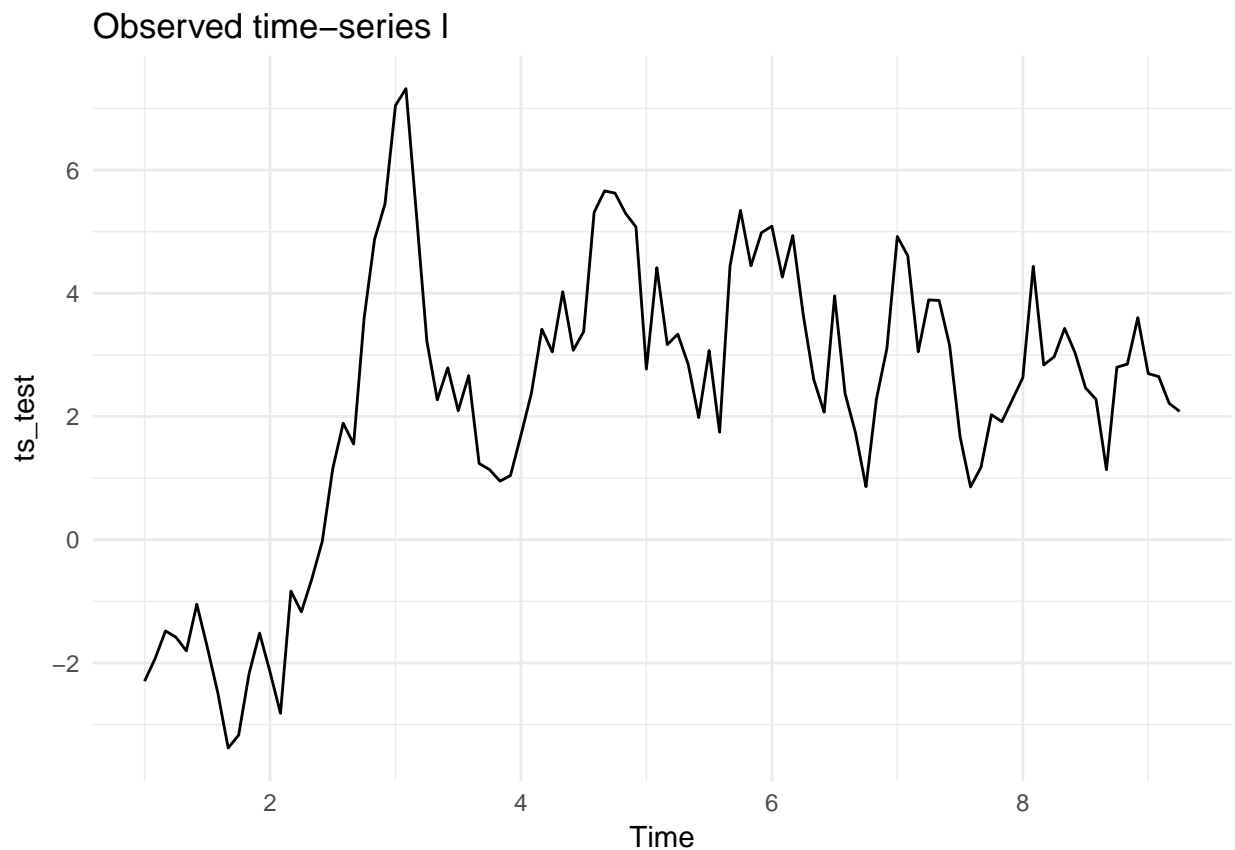
```
#no autocorrelation after detrending
#ggAcf(ts_test) +
# theme_minimal()
```

Part 1

The time-series plot suggests it is a strong autoregressive process where prior period levels are important for predicting $t+1$. The ACF provides the expected geometric decline in autocorrelation for an $AR(p)$ process. Further, the PACF shows that this series is an $AR(1)$ with $\phi = 0.9$.

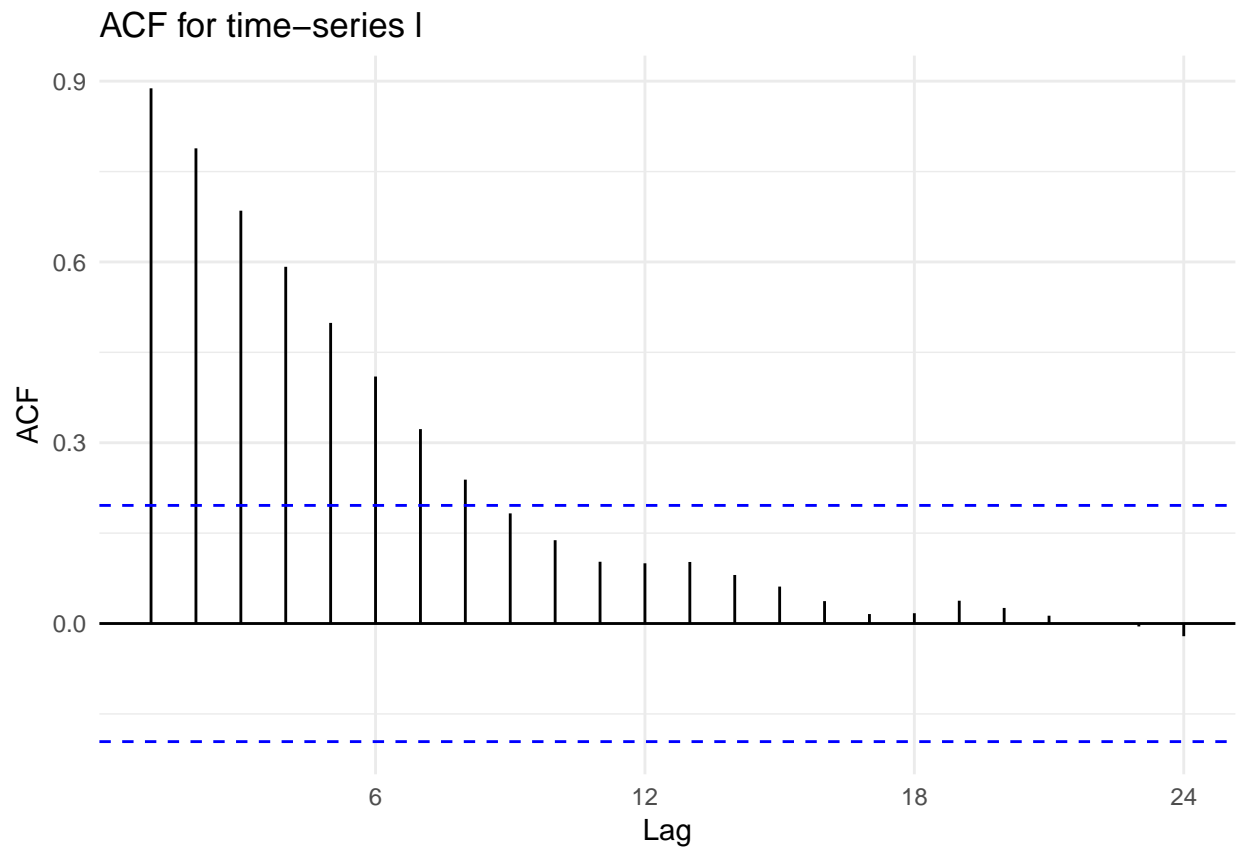
```
#pull time-series l
let <- "l"
ts_test <- ts(probl[[let]], frequency = 12)

#plot the time-series
autoplot(ts_test) +
  theme_minimal() +
  labs(title = "Observed time-series l")
```



```
#plot the decomposition
#autoplot(decompose(ts_test, type = "additive")) +
#  theme_minimal()

#plot the ACF
ggAcf(ts_test) +
  theme_minimal() +
  labs(title = "ACF for time-series l")
```



```
#plot the PACF  
#ggPacf(ts_test) +  
# theme_minimal() +  
# labs(title = "")
```


**** Multiple time-series characteristics ahead! ****

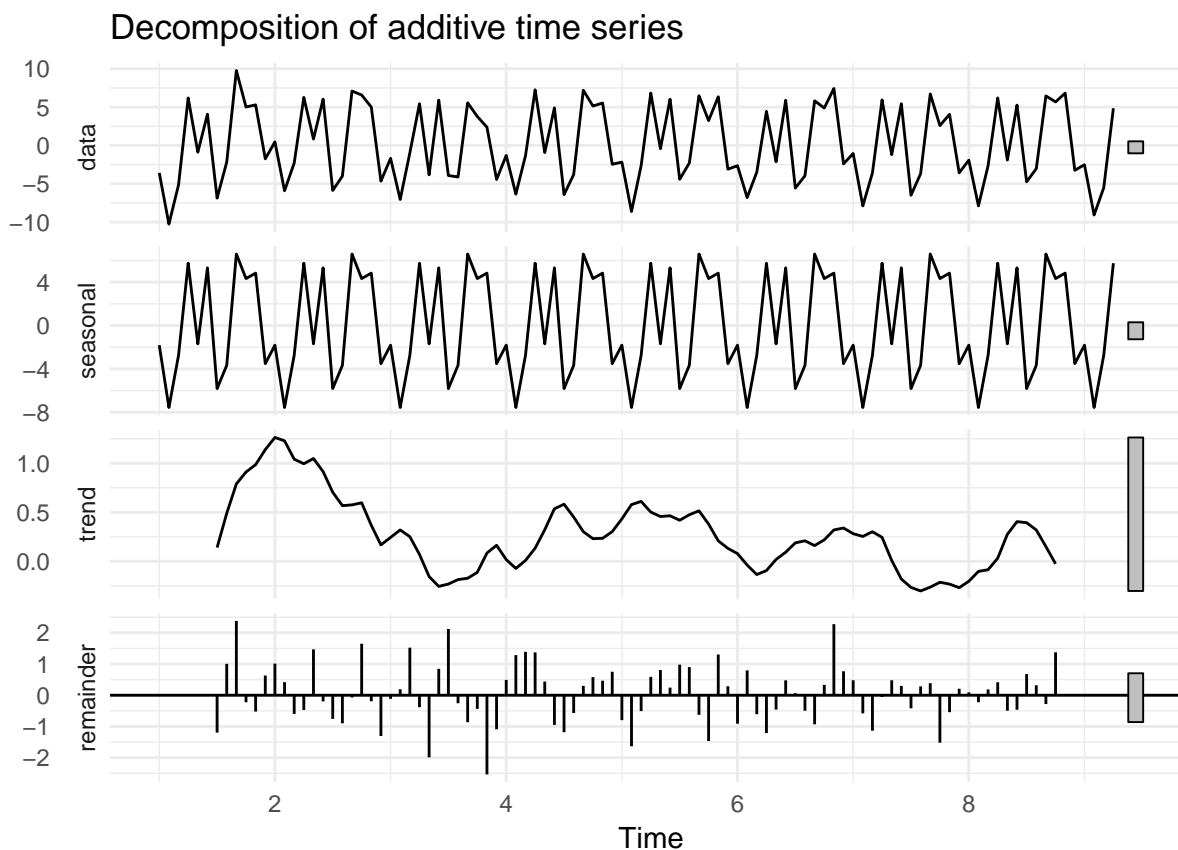
Part m

There seems to be a seasonal structure based on the time-series plot, but otherwise it's hard to make out much else about the DGP. The ACF has significant spikes at 12 and 24, and the decomposition model suggests a plausible seasonal pattern of higher levels throughout the middle of the year and lower levels in the winter that bottom out in February (this of course assumes that the time series starts at Jan year 1). The adjusted time-series still shows evidence of short-term autocorrelation, with a positive L1 according to the PACF and maybe a negative L2. An MA(2) process with $\psi_1 = 0.5, \psi_2 = -0.45$ characterizes the seasonally-adjusted time-series because there are shocks that last a couple periods (i.e. series stays above/below mean for a few periods), but the series does not drift much at all from the 0 mean overall (even in the T=100 sample).

```
#pull time-series m
let <- "m"
ts_test <- ts(prob1[[let]], frequency = 12)

#plot the time-series
#autoplot(ts_test) +
# theme_minimal()

autoplot(decompose(ts_test, type = "additive")) +
  theme_minimal()
```



```
#ggAcf(ts_test) +
# theme_minimal() +
```

```

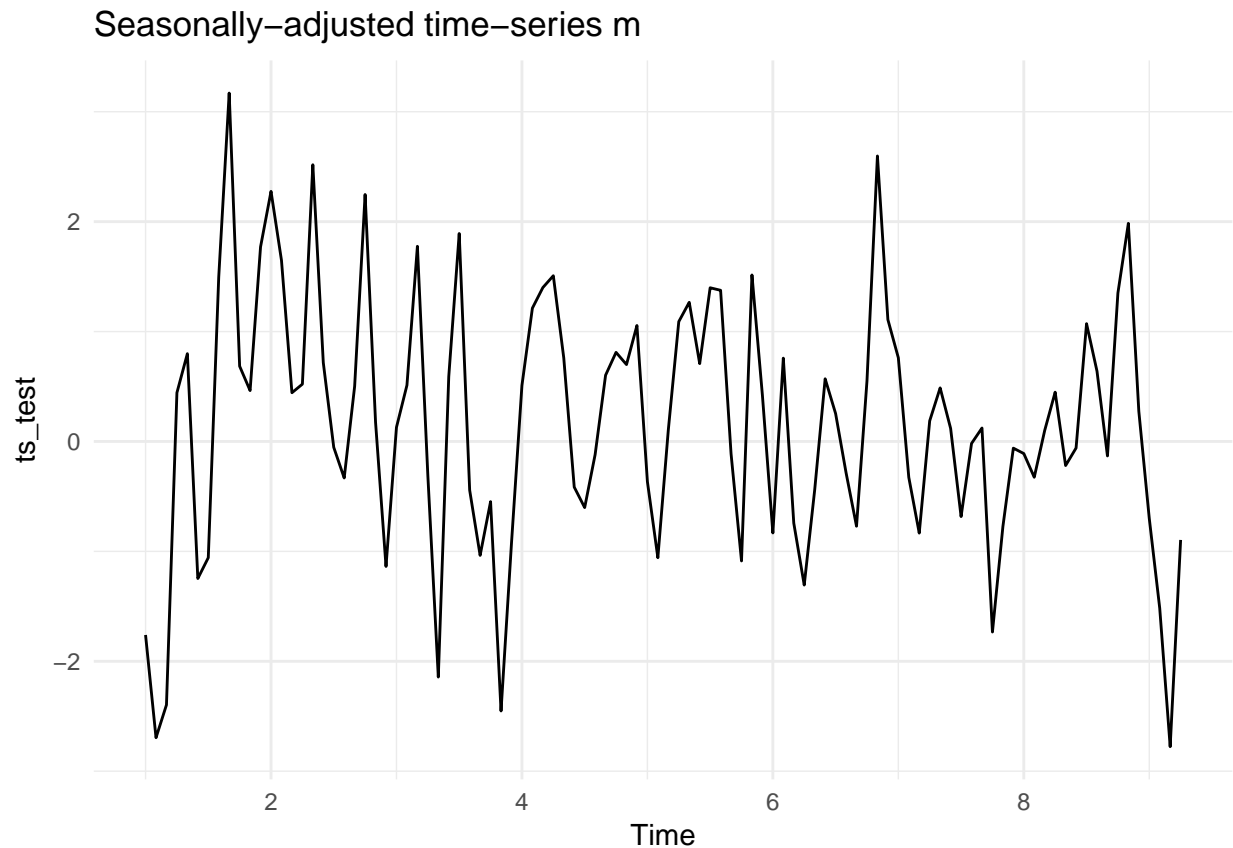
# labs(title = "")

#store decomposition model
decomp <- decompose(ts_test, type = "additive")

#seasonally adjust (y - seasonal mean for m)
ts_test <- ts_test - decomp$seasonal

#plot adjusted time-series
autoplot(ts_test) +
  labs(title = "Seasonally-adjusted time-series m") +
  theme_minimal()

```



```

#seasonally-adjusted ACF
#ggAcf(ts_test) +
# theme_minimal() +
# labs(title = "ACF for seasonally-adjusted time-series m")

#seasonally-adjusted PACF
#ggPacf(ts_test) +
# theme_minimal() +
# labs(title = "PACF for seasonally-adjusted time-series m")

#plot the simulated MA(2), psi_1 = 0.5, psi_2 = -.45
#autoplot(arima.sim(list(order = c(0, 0, 2), ma = c(0.5, -0.45)), n = 100)) +
# theme_minimal() +

```

```
# labs(title = TeX("Simulated MA(2),  $\psi_1 = 0.5, \psi_2 = -0.45$ "))
```

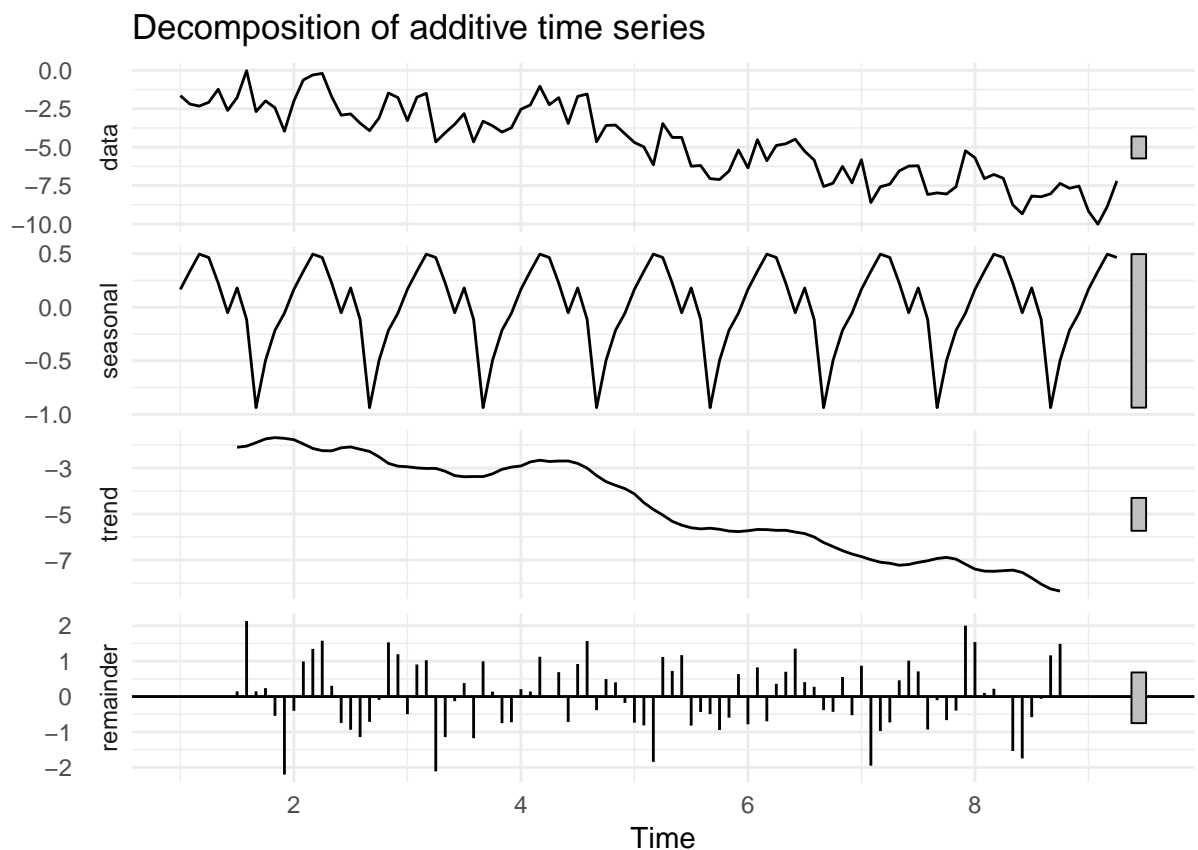
Part n

The time-series plot shows a clear negative trend over the course of the series, and this conclusion is supported by the decomposition model too where the trend component is negative. After detrending the time-series using the estimated monthly decrease from a linear model, the resulting series still shows some evidence of autoregression based on the ACF having a geometric decline in autocorrelation out to about L5. The PACF shows no significant autocorrelation past L1, so an AR(1) with $\phi = 0.45$ seems appropriate for this time-series.

```
#pull time-series n
let <- "n"
ts_test <- ts(prob1[[let]], frequency = 12)

#plot the time-series
#autoplot(ts_test) +
#  theme_minimal()

#plot the decomposition model
autoplot(decompose(ts_test, type = "additive")) +
  theme_minimal()
```

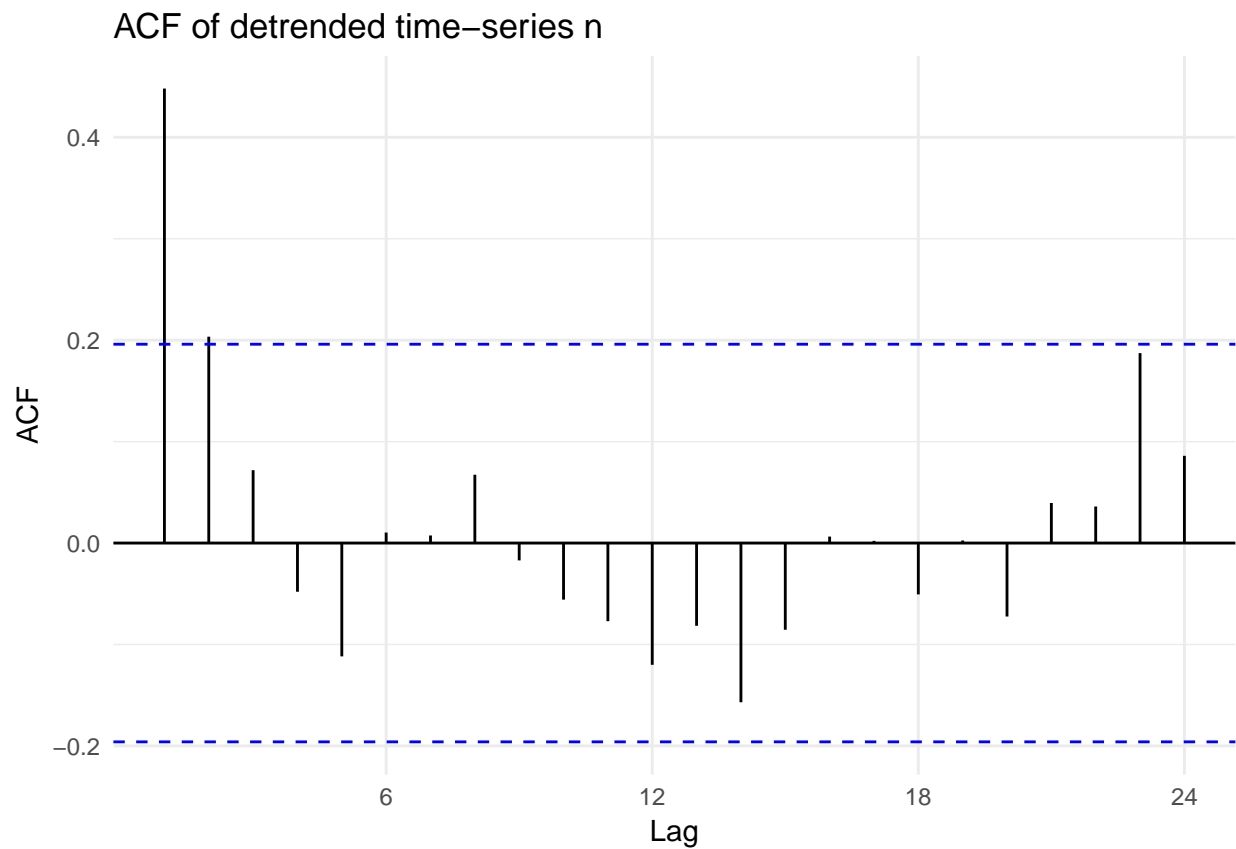


```
#looks like trend, let's estimate the monthly trend
time <- 1:100
lm_n <- lm(ts_test ~ time)
#summary(lm_n) #about -.075 increase per month

#subtract estimated trend from time-series
```

```
ts_test <- ts_test - lm_n$coefficients[2]*time

#still autocorrelation in the detrended data
ggAcf(ts_test) +
  theme_minimal() +
  labs(title = "ACF of detrended time-series n")
```



```
#looks like AR(1) is appropriate
#ggPacf(ts_test) +
# theme_minimal() +
# labs(title = "")
```

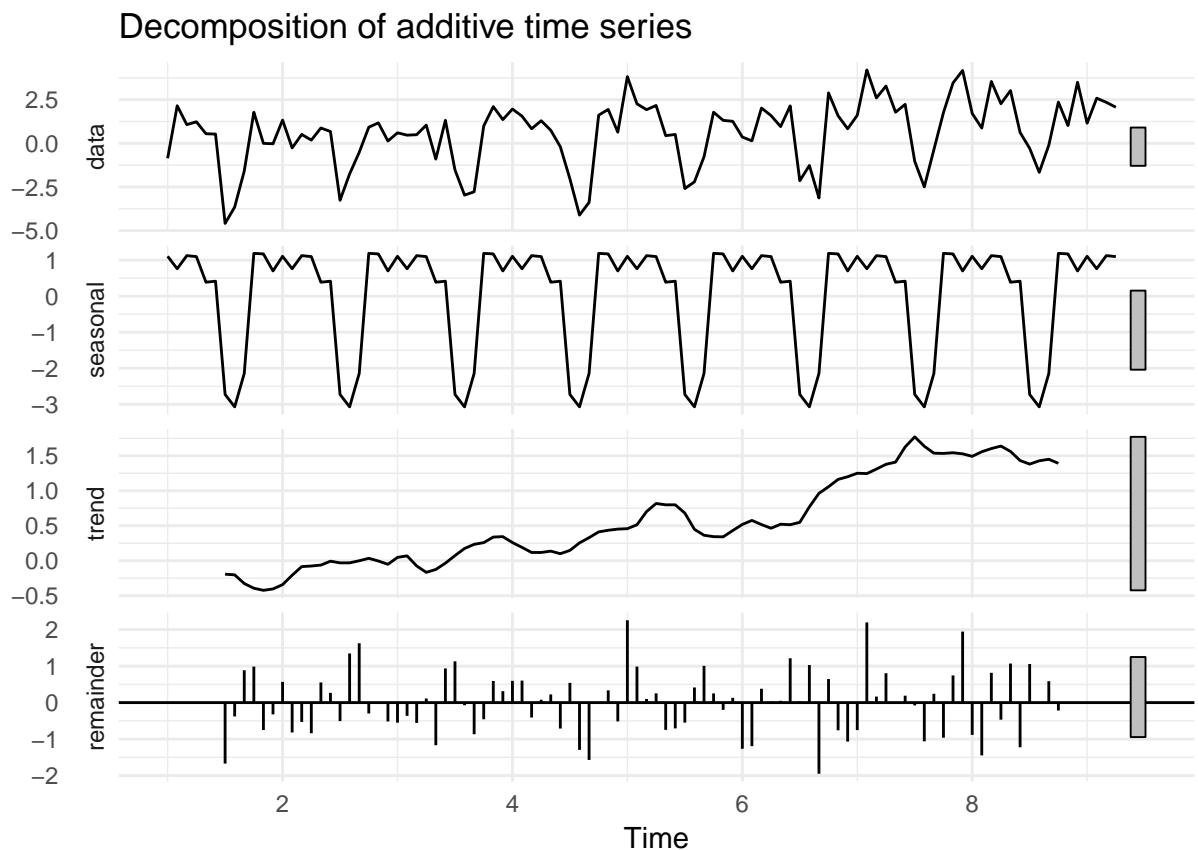
Part o

This time-series shows a weak positive trend and some seasonality according to the time-series plot and decomposition model. The decomposition shows that in two years cycles there is a pretty substantially dip in the levels in the middle of the second year, around which levels keep a pretty normal level. After removing both the trend and the seasonal cycle there does not appear to be any remaining autocorrelation in the series.

```
#pull time-series o
let <- "o"
ts_test <- ts(probi[[let]], frequency = 12)

#plot time-series
#autoplot(ts_test) +
#  theme_minimal()

#plot decomposition model
autoplot(decompose(ts_test, type = "additive")) +
  theme_minimal()
```



```
#store decomposition model
decomp <- decompose(ts_test, type = "additive")

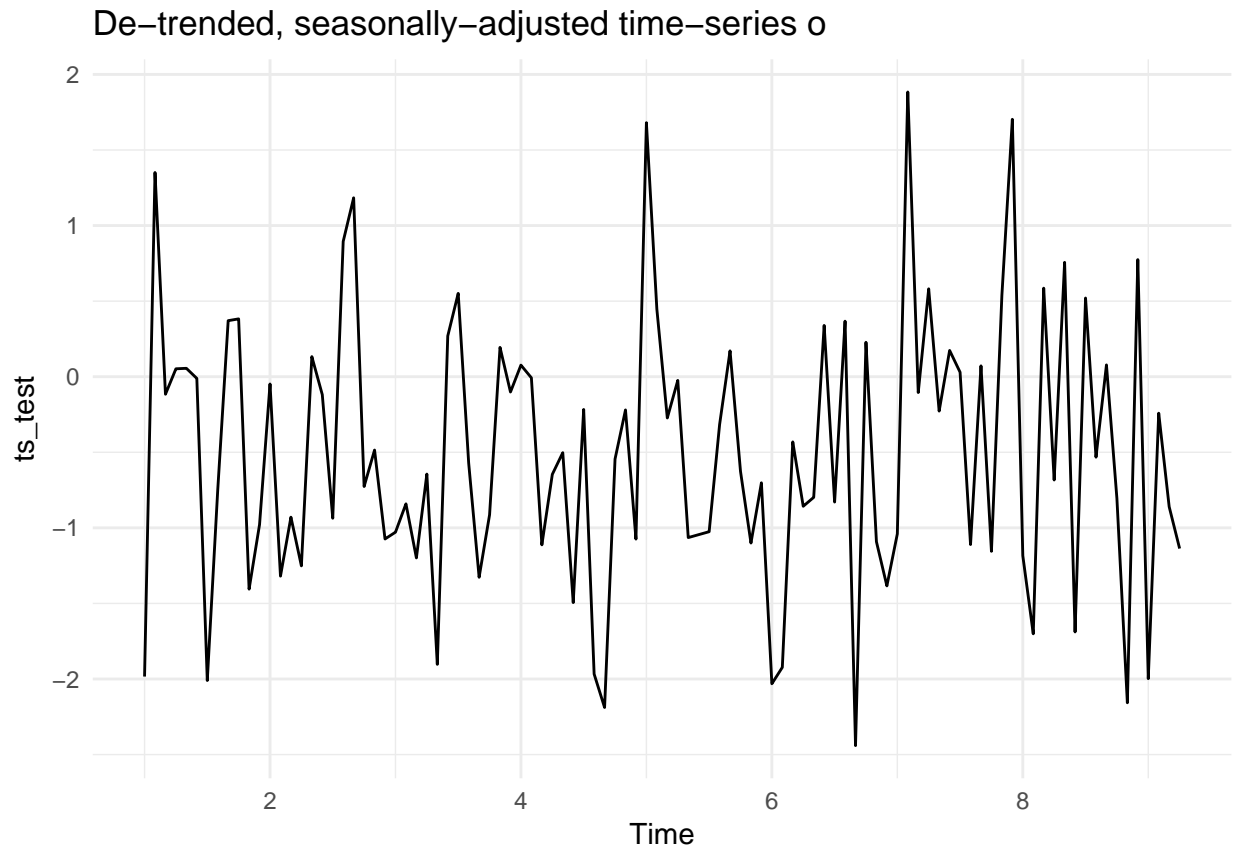
#seasonally adjust (y - seasonal mean for o)
ts_test <- ts_test - decomp$seasonal

#looks like trend, let's estimate the monthly trend
time <- 1:100
```

```
lm_o <- lm(ts_test ~ time)
#summary(lm_o) #about .02 increase per month

#subtract estimated trend from seasonally-adjusted time-series
ts_test <- ts_test - lm_o$coefficients[2]*time

#plot detrended, seasonally-adjusted time-series
autoplot(ts_test) +
  theme_minimal() +
  labs(title = "De-trended, seasonally-adjusted time-series o")
```



```
#no autocorrelation
#ggAcf(ts_test) +
# theme_minimal() +
# labs(title = "")
```

Part p

This time-series shows seasonality and trend in its plot according to time and in the decomposition model plots. Every years there is a cycle, peaking in the middle of the year in the cycle and bottoming out near the end. After removing the cycle and the trend, there is still significant autocorrelation, evidenced by the geometrically declining values on the ACF. The PACF has a significant spike of around 0.85 at L1, but otherwise shows that an AR(1) describes the de-trended, seasonally adjusted time-series.

```
#pull time-series p
let <- "p"
ts_test <- ts(prob1[[let]], frequency = 12)

#plot time-series
#autoplot(ts_test) +
#  theme_minimal()

#plot decomposition
#autoplot(decompose(ts_test, type = "additive")) +
#  theme_minimal()

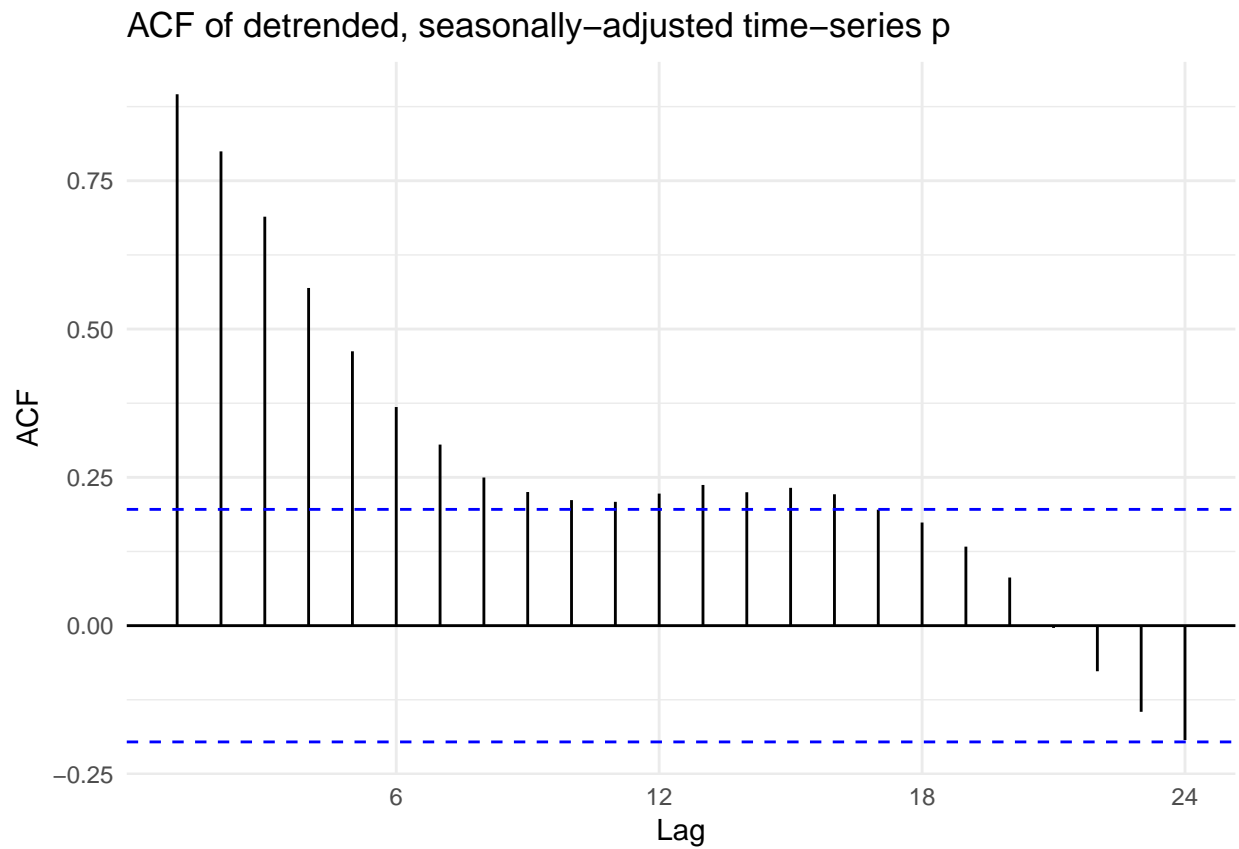
#store decomposition model
decomp <- decompose(ts_test, type = "additive")

#seasonally adjust (y - seasonal mean for p)
ts_test <- ts_test - decomp$seasonal

#estimate trend
time <- 1:100
lm_p <- lm(ts_test ~ time)

#subtract estimated trend from seasonally-adjusted time-series
ts_test <- ts_test - lm_p$coefficients[2]*time

#still strong residual autocorrelation
ggAcf(ts_test) +
  theme_minimal() +
  labs(title = "ACF of detrended, seasonally-adjusted time-series p")
```

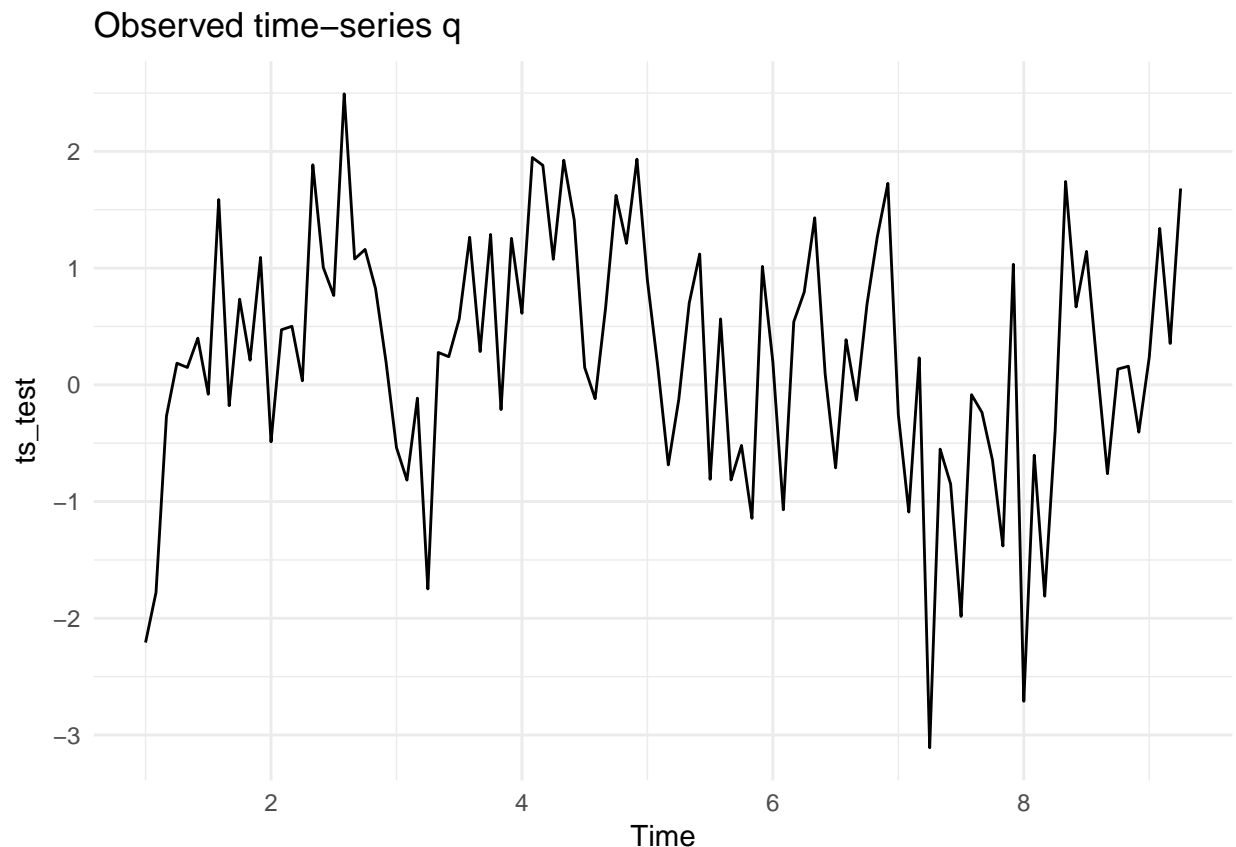
```
#suggests AR(1), phi = 0.95
#ggPacf(ts_test) +
# theme_minimal() +
# labs(title = "")
```

Part q

This time-series does shows a weak seasonal pattern, with a rough pattern of increase throughout the year but low January-February values. After adjusting the series by subtracting the estimated monthly means, there is a pretty clear pattern of autocorrelation in the ACF plot. The PACF plot suggests that the order of the AR(p) process is 2, with mild positive effects for both coefficients. I estimate the coefficient magnitudes for the AR terms to be about 0.4 and 0.2, respectively.

```
#pull time-series q
let <- "q"
ts_test <- ts(probl[[let]], frequency = 12)

#plot time-series
autoplot(ts_test) +
  theme_minimal() +
  labs(title = "Observed time-series q")
```



```
#plot decomposition
#autoplot(decompose(ts_test, type = "additive")) +
# theme_minimal()

#store decomposition model
decomp <- decompose(ts_test, type = "additive")

#seasonally adjust (y - seasonal mean for q)
ts_test <- ts_test - decomp$seasonal
```

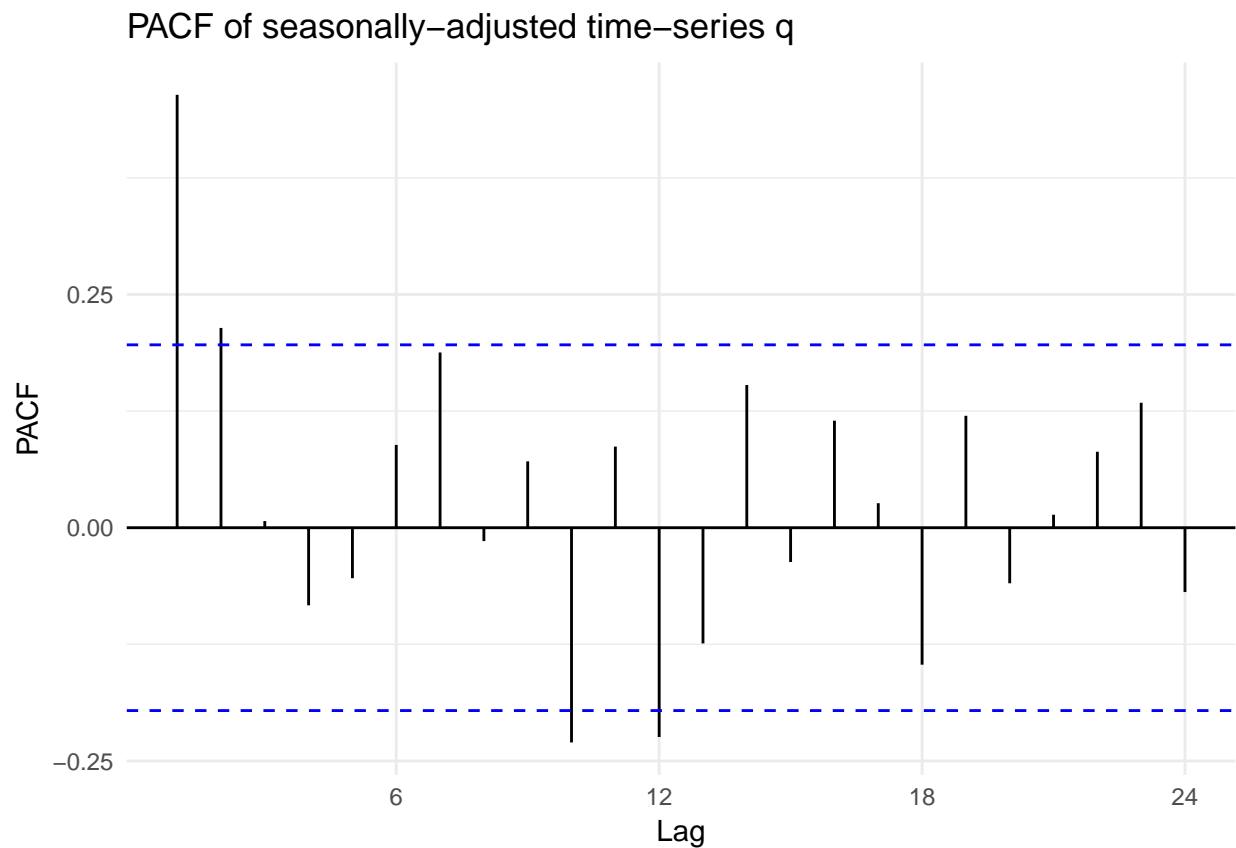
```

#plot seasonally-adjusted time-series
#autoplot(ts_test) +
# theme_minimal()

#dies off L6
#ggAcf(ts_test) +
# theme_minimal() +
# labs(title = "")

#only two, both positive
ggPacf(ts_test) +
  theme_minimal() +
  labs(title = "PACF of seasonally-adjusted time-series q")

```



```

#plot simulated AR(2) with phi_1 =.4, phi_2 =.2
#autoplot(arima.sim(list(order = c(2, 0, 0), ar= c(0.4, 0.2)), n = 100)) +
# labs(title = TeX("Simulated AR(2),  $\phi_1 = 0.4$ ,  $\phi_2 = 0.2$ ")) +
# theme_minimal()

```

Part r

The time-series plot and decomposition model both seem to indicate a positive deterministic trend over the span of the time-series. After removing the trend, there is substantial autocorrelation at L1, but nothing systematic beyond that. At this point, the non-geometric decline in autocorrelation suggests a moving-average process. The PACF mostly corroborates that a MA(3) would be appropriate, with a moderate ψ_1 of about 0.5 and two small higher-order terms with values around .20.

```
#pull time-series r
let <- "r"
ts_test <- ts(probl[[let]], frequency = 12)

#plot time-series
#autoplot(ts_test) +
#  theme_minimal()

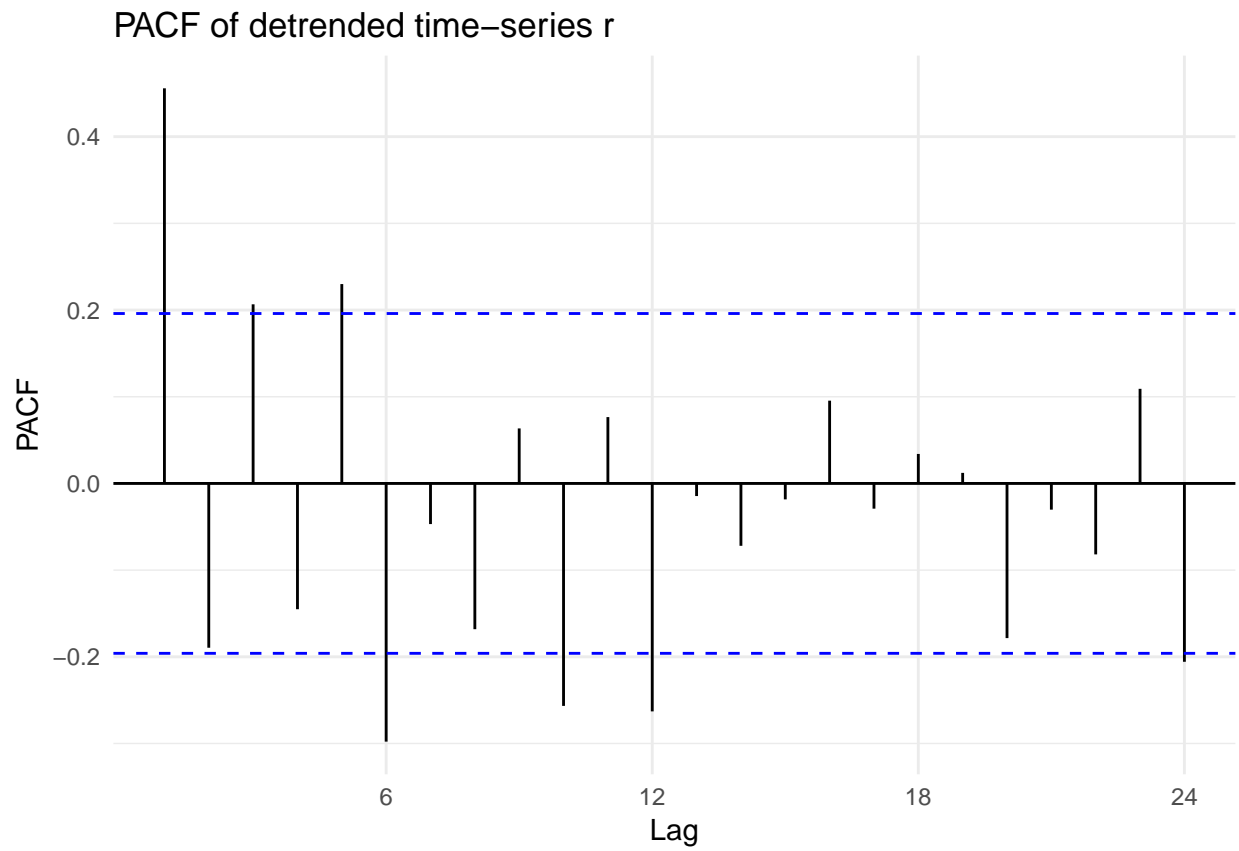
#autoplot(decompose(ts_test, type = "additive")) +
#  theme_minimal()

#estimate trend
time <- 1:100
lm_r <- lm(ts_test ~ time)

#subtract estimated trend from seasonally-adjusted time-series
ts_test <- ts_test - lm_r$coefficients[2]*time

#ggAcf(ts_test) +
#  theme_minimal() +
#  labs(title = "ACF of detrended time-series r")

ggPacf(ts_test) +
  theme_minimal() +
  labs(title = "PACF of detrended time-series r")
```



Problem 2

Part s

- at 20** AR(1) with $\phi = .75$. It seems like it could be stationary with a mean around 5 (knowing nothing else about the DGP), so it'd be a little early to rule non-stationary. The Dickey-Fuller test said non-stationary, though it could just be because the data length was short.
- at 100** At this point I'd say AR(1) with $\phi = 1.0$ because the ACF/PACF will be unreliable in the presence of non-stationarity. I am more certain about ruling the ACF/PACF as unreliable because the time-series shows substantial autocorrelation and no apparent reversion to a mean over 100 observations. The Dickey-Fuller test was still not significant with 100 observations, so more evidence in favor of non-stationarity.
- overall** It was not immediate clear before but it became very clear that this is a random-walk, i.e. AR(1) with $\phi = 1.0$ using the full time-series. The ACF is so close to 1.0 that it should raise eyebrows alone, and the test of stationarity still points in favor of the opposite conclusion.

Part t

- at 20** This subset looks stationary with a mean of -1, probably an AR(1) with $\phi = 0.25$ or so. Though the Dickey-Fuller does suggest non-stationarity, it seems difficult to bank on that with relatively short data.
- at 100** This subset still looks stationary with a mean of -1, an AR(1) with $\phi = 0.75$. So more clearly autoregressive.
- overall** Definitely stationary although the DGP mean looks to be 0 not -1, AR(1) with $\phi = 0.8$ seems appropriate. I am more certain to conclude it now since there are two clear signs (i.e. ts plot and Dickey-Fuller).

Part u

- at 20** The time-series plot suggests non-stationarity, but this is admittedly a tough call since it makes some strong bounces away from 0 (which is what the mean looks like it'd be if stationary). I'd say AR(1) with $\phi = 1.0$ right now. The ACFs show very strong autocorrelation despite short data, and would be unreliable if there is a unit root.
- at 100** Harder to say now, the first subset might have just been part of a cyclical pattern or might instead be random walks far in the negative direction that moved positive and stayed there for most of the other 80 observations. ACF/PACF say AR(1) with $\phi = 0.95$ and mean of 0, but at this point I think I'd still conclude it is AR(1) with $\phi = 1.0$
- overall** Over the 1000 observations, it does turn out to just be a strong AR(1) with $\phi = .95$ and mean of 0. It is clearer that some of the variability (increase with some cyclical decrease) was part of a seasonal structure.

Part v

- at 20** With only 20 observations it is very hard to judge. It could be non-stationary AR(1) with $\phi = 1.0$, but at the same time it could have a mean of 0 and just be bouncing away temporarily. Not knowing anything else about the data I would be cautious about taking the ACF at face value given the observed time-series

at 100 Now it looks more likely to be non-stationary, though ACF/PACF/dickey-fuller say stationary with *really* strong AR(1) with 0.95 phi. The strong deviations from 0 (seemingly the mean) suggest that the time-series is non-stationary and likely an AR(1) process

overall With the 1000 observations, it's clearly non-stationary

Part w

at 20 It looks stationary, like the prior period has some importance/moderate autocorrelation probably AR(1) with $\phi = .55$

at 100 The 100 observation subset looks less stationary because the series is spending long time stretches away from the mean (which now seems like it'd be zero). Though there are cycles of increasing amplitude over the 100 periods, I think I'd conclude AR(1) with $\phi = 1.0$ based on the ACF/PACF and Dickey-Fuller test. Though the PACF shows there could be a second-order term, this could also be unreliability from non-stationarity or a significant L2 just due to chance (i.e. a more parsimonious explanation seems merited given the evidence).

overall It looks non-stationary based on the overall plot because there's no clear structure of how the prior day(s) or cycles matter for $t+1$ and there's not a clear pattern of mean reversion either (though the series does pass 0 a few times). This conclusion is supported by Dickey-Fuller test. The ACF/PACF are so close to 1.0, it would seem tenuous to accept them on face value here.

Problem 3

My proposed research design is to model a quarterly panel of census tract neighborhood rent levels for the Seattle area (maybe King County as a whole). For about the past year, a team of collaborators and I have been scraping Craigslist to generate a database of unique rental housing listings. The outcomes will accordingly be the neighborhood median estimates for rent asked by specific bedroom sizes. We will have 4 periods for T (Qtr 2,3,4 2017, Q1 2018). I intend to first model differences in these distributions using spatio-temporal smoothing, i.e. no covariates and just subject-specific adjustments for spatial structure, temporal structure and their interaction. After generating spatially-and-temporally smoothed estimates of neighborhood rent distributions for the metropolitan area, I will compare these with existing data sources to assess agreement between model-based CL estimates and existing (largely for-profit) data sources (see after this problem for a few examples of existing comparisons). I plan to use INLA to estimate these Bayesian models. This qualifies as a prediction problem more than an explanation problem as currently devised, so I'd like to demonstrate agreement but also investigate the places where there is substantial disagreement between measures. The data are somewhat unusual in their sampling strategy, i.e. scraping web data, rather than surveying landlords about what they charge their tenants. Figure 1 shows a time-series heatmap for our unique listings collected per day for the Seattle area.

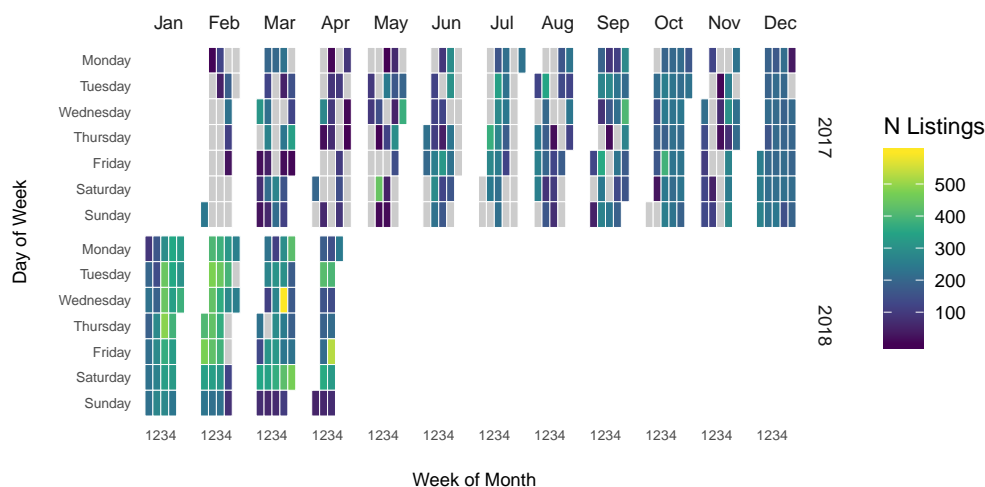


Figure 1: Time-series heatmap of Craigslist rental housing listings

While we have increasing confidence that we are sampling all of what's on Craigslist at our current stage, earlier in our project we no doubt were missing some listings since our scraper was less robust to page changes and we had more downtime in general. As such, we have some missing data when looking at the city overall because the scraper was down, and when looking at neighborhoods there are some neighborhoods where we have much fewer listings. Some of this is potentially due to process, but we also find evidence of variations in neighborhood listing activity. Spatial and non-spatial Poisson models find that the listings generally follow characteristics of the housing stock (i.e. more units -> more listings, more owner-occupied HU -> fewer listings) but there's a strong pattern of listing activity being negatively associated with neighborhood poverty and to a lesser degree the size of the foreign-born population. So while there is not missing data in many tracts, there do seem to be substantial heterogeneity in the use of Craigslist versus using word-of-mouth or physical advertisements (the theoretical housing search mechanisms in these areas). The goal of space/time smoothing is to adjust our estimates to most efficiently use all

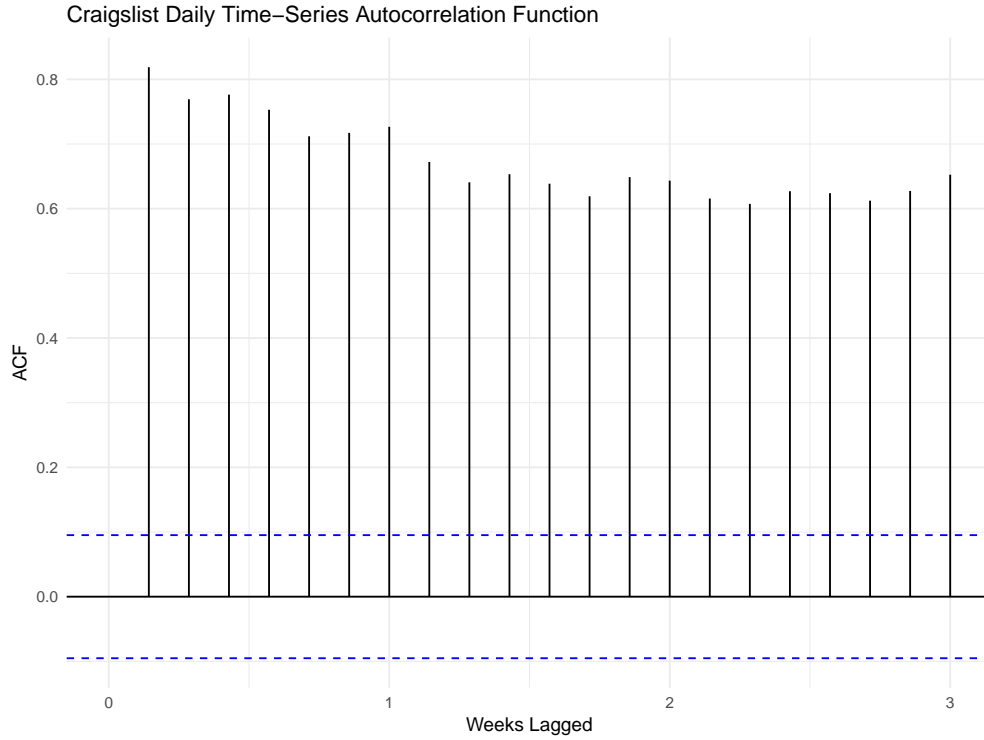


Figure 2: ACF of Craigslist median rent time-series

I believe that there is autoregression in the panel data given my investigation of the city-wide median rent level as a daily time-series. Figure 2 shows the ACF for this citywide median daily time-series, and you can see a strong autoregressive process with some weekly seasonality. I have not yet looked at whether this is the case for the panel of Seattle neighborhoods, but expect even with aggregating to a larger T there will be autocorrelation (though the weekly cycles will be averaged out.). It is hard to imagine a scenario where a future quarter's rent value would not still be highly predicted by the prior quarter's level or shock. The focal GM assumptions that would be violated in running a linear regression are serial correlation and heteroskedasticity since there's both temporal and spatial structure present in the panel data. Treating observations of the same unit or of different units as having the same error around them ignores the serial correlation within units and the spatial configuration between units.

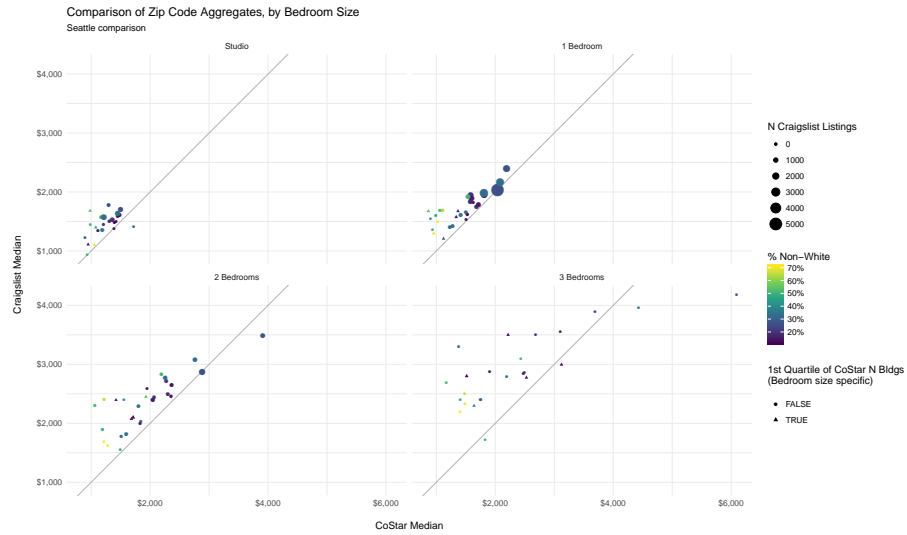


Figure 3: Craigslist / Apartments.com-CoStar comparison (zipcode level)

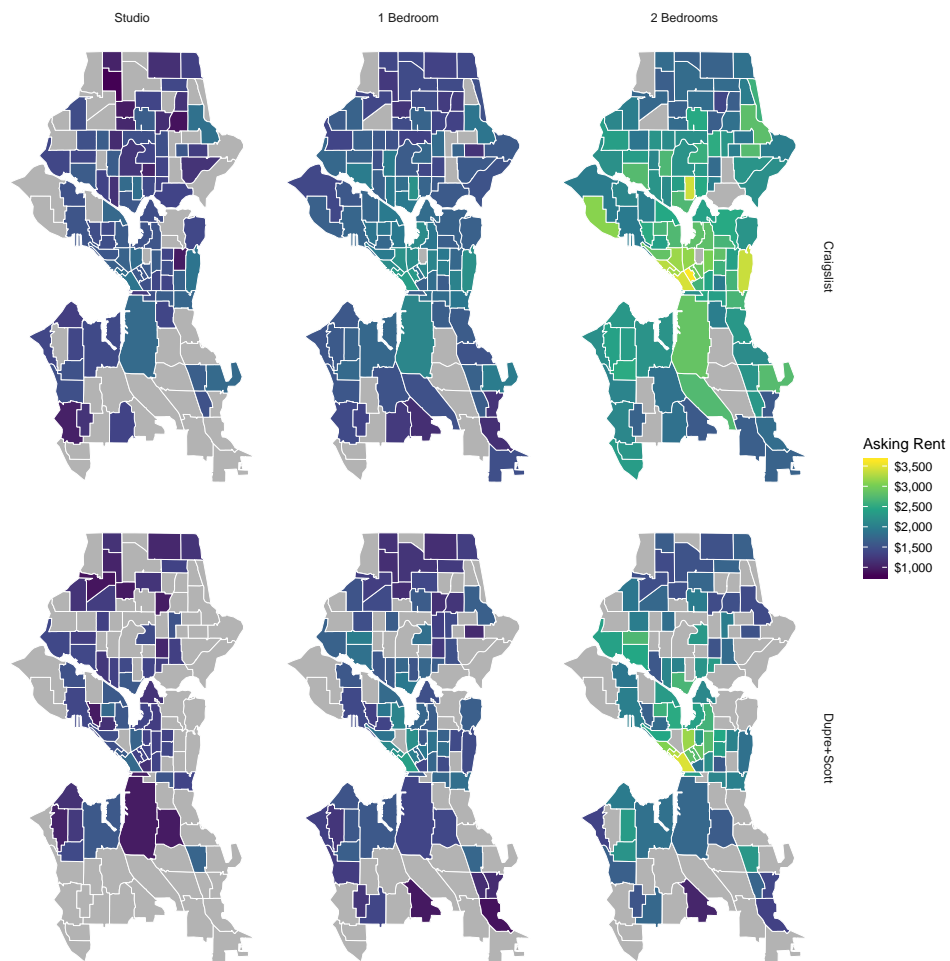


Figure 4: Craigslist / Dupre+Scott comparison (tract level)