

Time Series Homework 1

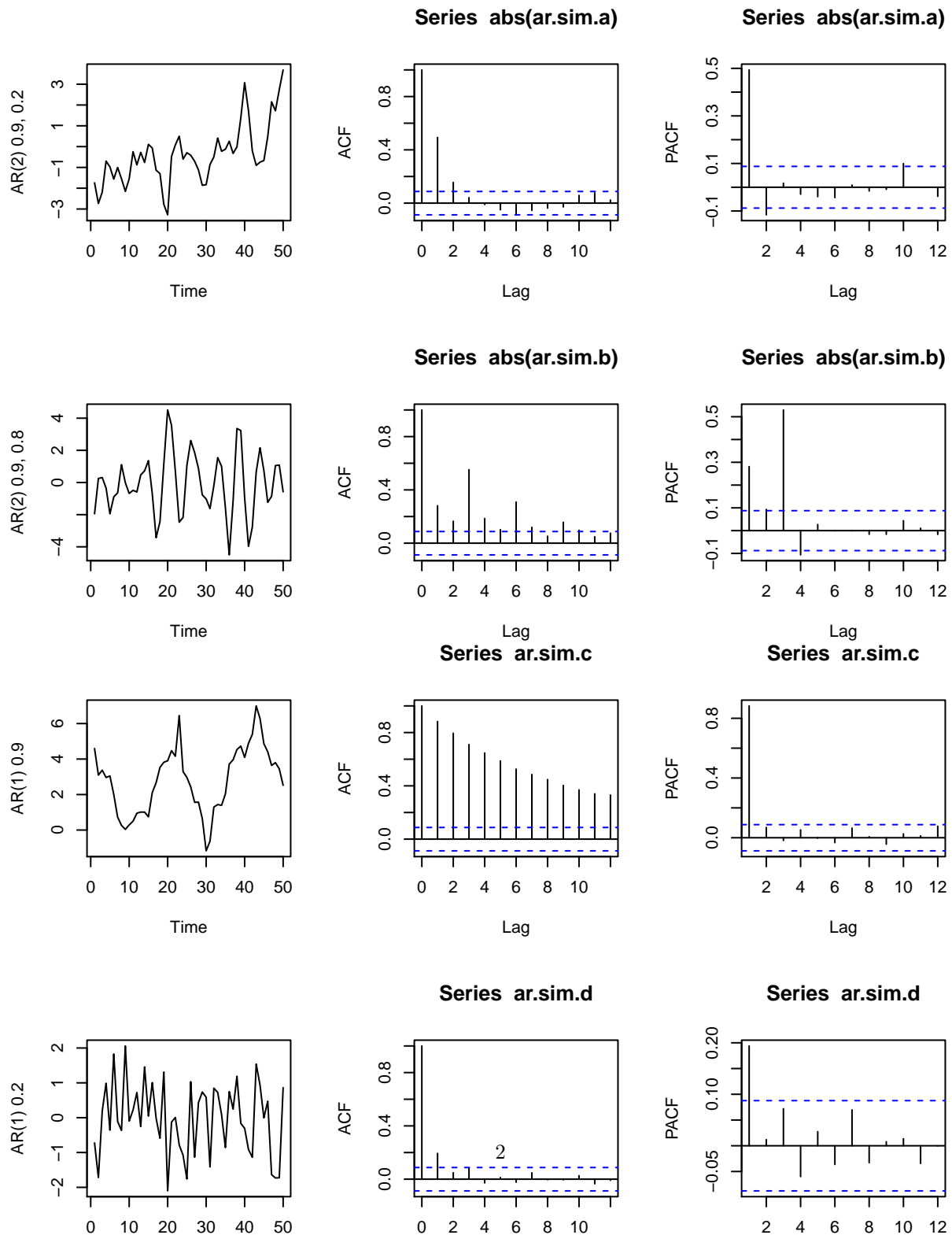
Kylie Taylor

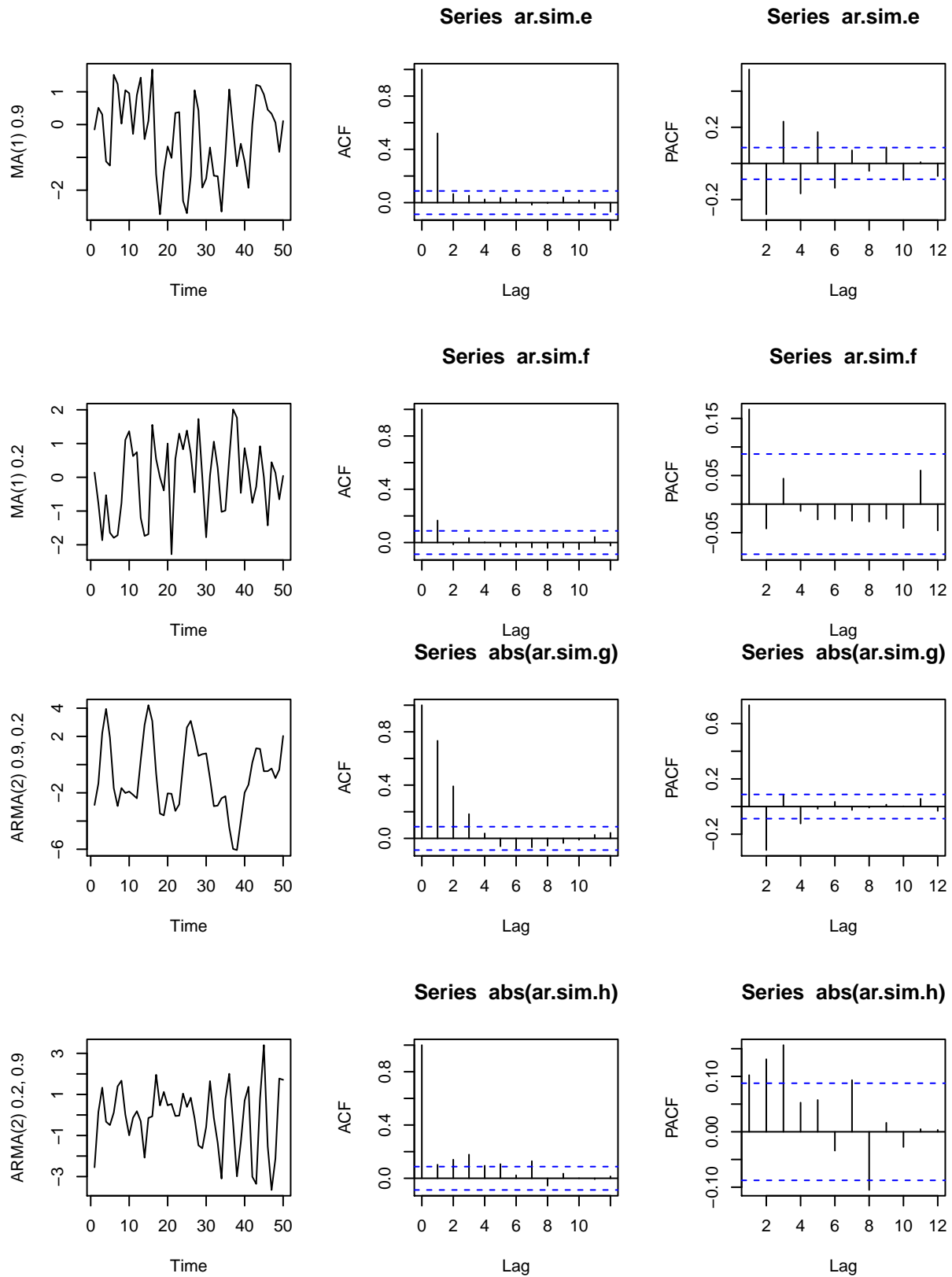
2/5/2019

1.

Part A: Compute ACFs and PACFs, make comparisons and comment on the following:

would you be able to assign the right model to the generating series if you were not aware of the generating process? How, why?





The output from the eight different time series models is above. Simply by looking at the ACF's and PACF's, I would not be able to assign the correct model to the generating series if I was not aware of the generating process. The ACF's and PACF's appear to be equally as random as the series themselves. For example, take

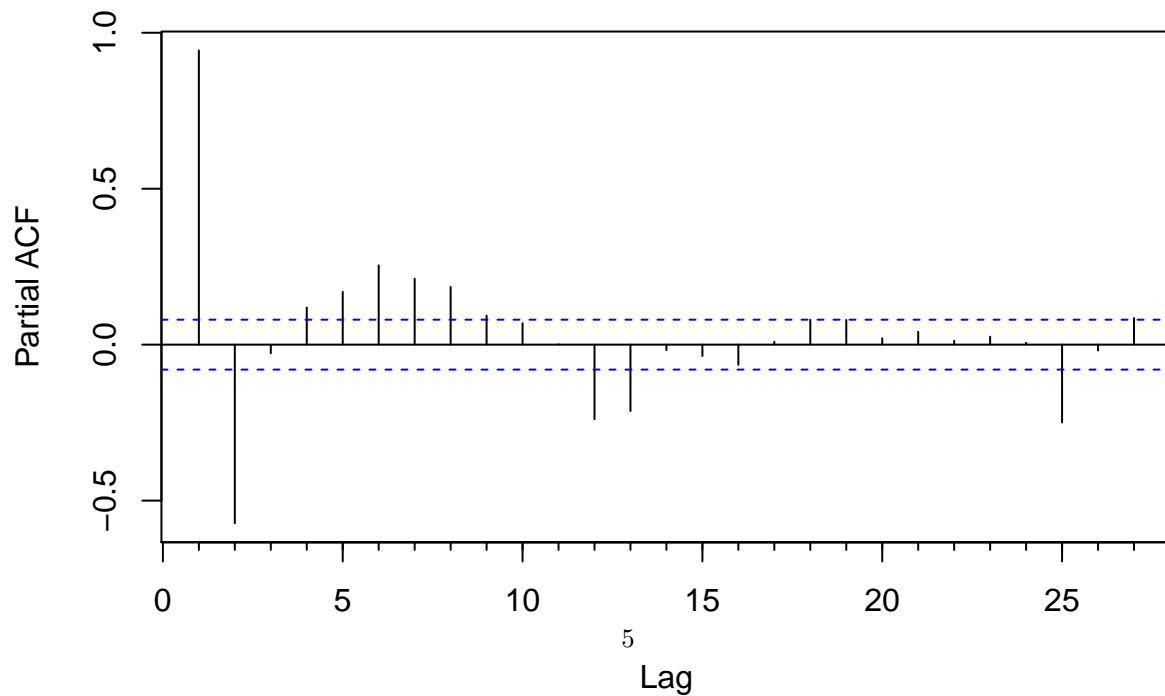
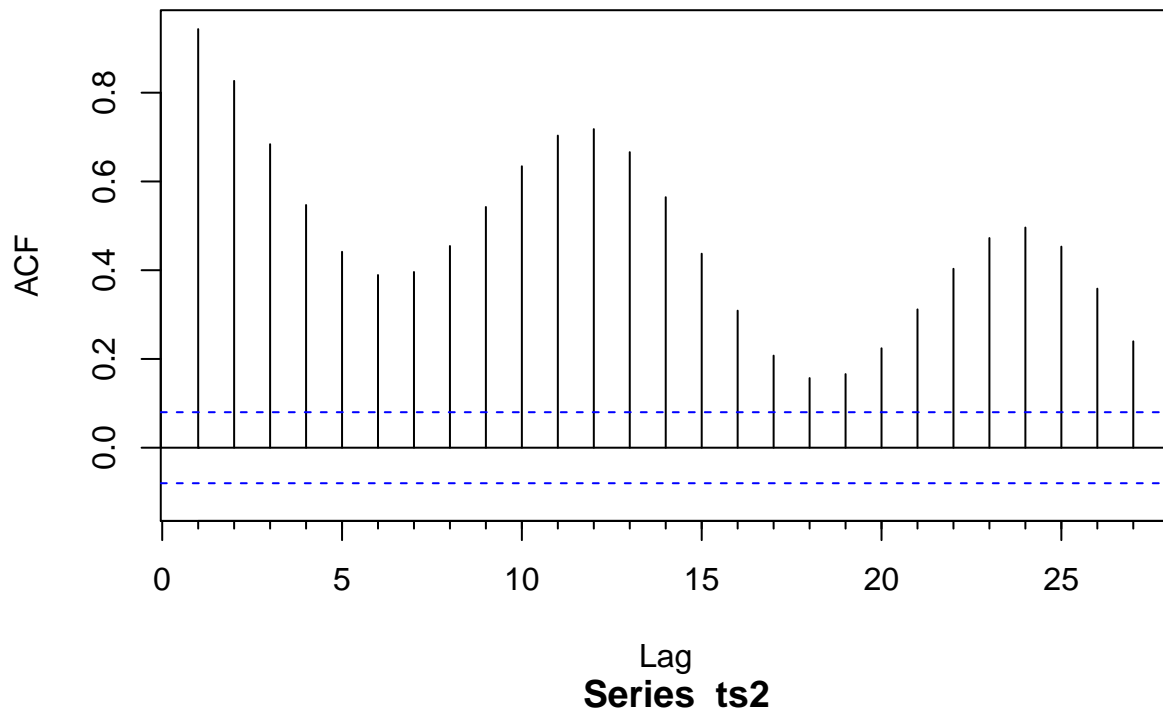
the AR(1) with coefficient 0.2 and the MA(1) with the coefficient 0.2, the PACF's look very similar, even though the underlying model of their respective PACF's are very different. Another example is the ACF's for the AR(2) with coefficients 0.9 and 0.2, AR(1) with coefficient 0.9, and the ARMA(2) with coefficients 0.9 and 0.2, the plot of the ACF's all appear to follow the similar decaying correlations, that I would not be able to discern between if I did not know the underlying process.

The reason it is hard (almost impossible) to determine the generating process by simply looking at the ACF's and PACF's is because the generating process is inherently random (across time), therefore every process will not follow the same forms.

Part B :

Use a data series of your choice to create ACF's and PACF's. Comment on the persistence of the series, and argue on a model that you would choose for modeling this series. Remove trend and take first differences if necessary.

Series ts2



The data set I picked was the monthly water level of Lake Erie from 1921 to 1970. I detrended the data and generated the ACF's and PACF's. Observing the ACF, I would suggest a ARMA(3) model, with lags of 1, 2 and 12 periods. The ACF appears to follow a cyclical and decaying pattern. This would make sense because I would expect the level of the lake to be similar at the same time each year, also similar to the month before. The lake is likely not to dramatically increase or decrease in water levels within the span of a month. The PACF verifies my thought of including three lags for 1 month, 2 months and 12 months, because we see peaks in the PACF at those periods.

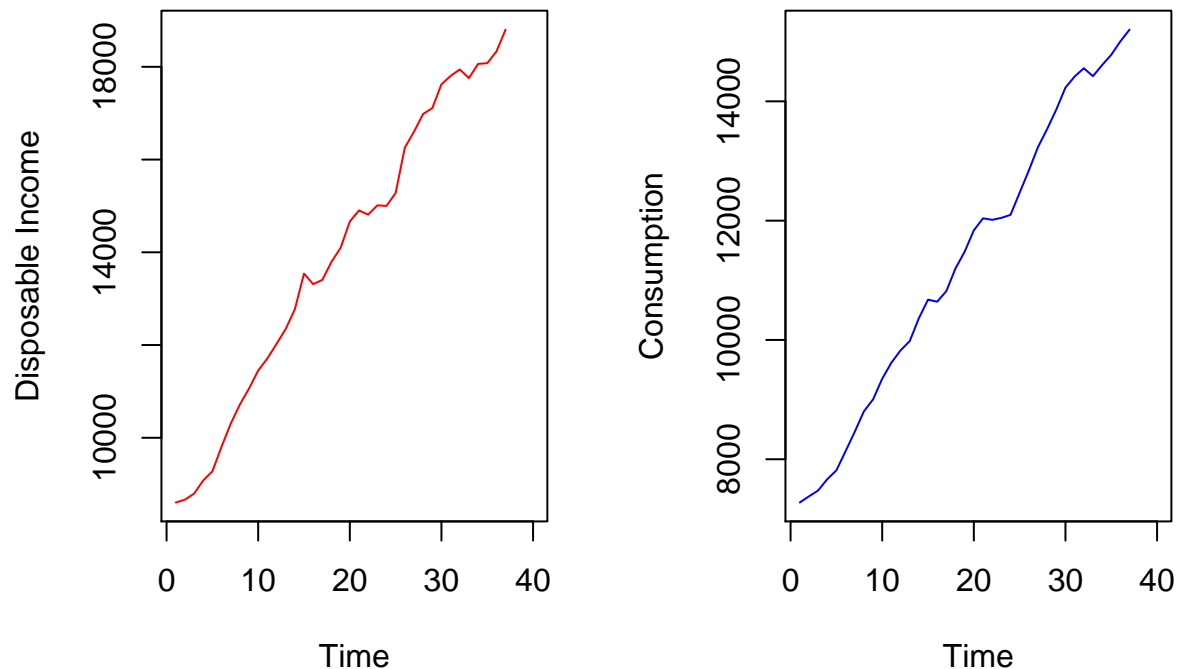
2.

a. Run a simple regression of real per capita consumption on real per capita income. Discuss your results; are these results reliable? Comment also on your usual goodness of fit statistics.

```
##
## =====
##                      Dependent variable:
##                      -----
##                      c
## -----
## y                      0.779***
##                      (0.007)
##
## Constant              463.180***
##                      (98.791)
##
## -----
## Observations           37
## R2                     0.997
## Adjusted R2            0.997
## Residual Std. Error    133.092 (df = 35)
## F Statistic            12,720.570*** (df = 1; 35)
## =====
## Note:                  *p<0.1; **p<0.05; ***p<0.01
```

The results reveal that for each additional unit of real per capita income, the average per capita consumption is expected to increase by 0.779 units, holding all else fixed. The R^2 of this model is 0.997, which reveals that the variation in income explains 99.7% of the variation in consumption. These results appear to be reliable to me. The F-stat of 12,720.57 reveals that the fit of this model is very good.

b. Plot the two series; what type of trend do you think they have?



These two series obviously have a positive trend, and they appear to follow the same trend.

c. Run the same regression but include the appropriate trend. How do the results change? Comment also on your usual goodness of fit statistics.

```
##
## =====
##                               Dependent variable:
##                               -----
##                               c
## -----
## y                               0.508***
##                               (0.046)
##
## Trend                           80.912***
##                               (13.654)
##
## Constant                        2,707.839***
##                               (385.254)
##
## -----
## Observations                     37
## R2                               0.999
## Adjusted R2                      0.999
## Residual Std. Error      94.709 (df = 34)
## F Statistic      12,577.640*** (df = 2; 34)
```

```
## =====
## Note:          *p<0.1; **p<0.05; ***p<0.01
```

I detrended both the per capita disposable income and the per capita consumption in the model above. The coefficient becomes larger, now 0.508, is still significant, with an R^2 of 0.998. The coefficient on the trend variable is significant at the 10% level. The F-stat reveals that the fit of this model is still very good.

d. Run regressions of consumption on a constant and the appropriate model of trend and of income on a constant and the appropriate model of trend and save the residuals; those residuals can be thought of as the de-trended series. Run a regression of the de-trended consumption on the de-trended income. How these results relate to your results in part c? Comment also on your usual goodness of fit statistics.

```
##
## =====
##               Dependent variable:
##               -----
##               resid.c
## -----
## resid.y              0.779***
##                   (0.007)
##
## Constant              -0.000
##                   (21.880)
##
## -----
## Observations              37
## R2                        0.997
## Adjusted R2              0.997
## Residual Std. Error    133.092 (df = 35)
## F Statistic            12,720.570*** (df = 1; 35)
## =====
## Note:          *p<0.1; **p<0.05; ***p<0.01
```

The results from this regression of “de-trended” variables, has a coefficient of 0.779, which is still significant. The major difference is that the intercept is now zero, compared to a very large, negative intercept above. The R^2 remains high and the goodness of fit is still very acceptable.

e. Compute the growth rate of both series (using the change in logarithms) and run a regression of the growth rate of consumption on the growth rate of income. Comment on your results.

```
##
## =====
##               Dependent variable:
##               -----
```



```

##                                lc
## -----
## ly                                0.944***
##                                (0.009)
##
## Constant                        0.329***
##                                (0.082)
##
## -----
## Observations                     37
## R2                               0.997
## Adjusted R2                     0.997
## Residual Std. Error      0.013 (df = 35)
## F Statistic      11,959.610*** (df = 1; 35)
## =====
## Note:          *p<0.1; **p<0.05; ***p<0.01

```

The coefficient of de-trended consumption is 0.944. This makes perfect sense because we would expect a roughly proportional percentage change in per capita income for any change in per capita consumption. This again is reflected by a very high R^2 and very large F-stat.

f. Run the same regression as part e., but now add 4 lags of income growth on the right hand side. Compute the contemporaneous and the long run effect of income growth on consumption, and graph it, together with 95% standard error bands.

```

##
## =====
##                                Dependent variable:
##                                -----
##                                lc
## -----
## ly                                0.719***
##                                (0.026)
##
## L(Lag, 4)                        0.005***
##                                (0.001)
##
## Constant                        2.373***
##                                (0.232)
##
## -----
## Observations                     37
## R2                               0.999
## Adjusted R2                     0.999
## Residual Std. Error      0.007 (df = 34)
## F Statistic      19,635.840*** (df = 2; 34)
## =====
## Note:          *p<0.1; **p<0.05; ***p<0.01

```

3.

Part A: Discuss small sample and large sample assumptions and OLS properties for Time Series; how do these assumptions compare to cross sectional data? Why do we need to rely more on the large sample assumptions for Time Series compared to Cross Sectional data?

Small (finite) sample assumptions: Linear in parameters, no perfect collinearity, strict exogeneity, homoskedasticity, no serial correlation, and normality.

Large (asymptotic) sample: Linearity and Weak dependence, no perfect collinearity, contemporaneous exogeneity, homoskedasticity, and no serial correlation.

The major difference between small and large sample assumptions is that large sample assumptions require linearity and weak dependence, whereas small sample assumptions require linear parameters. Small sample assumptions require strict exogeneity, which implies contemporaneous exogeneity, whereas large sample assumptions only require contemporaneous exogeneity.

These assumptions vary from cross-sectional data because with cross-sectional data we make the assumption that the errors are i.i.d. $\text{Normal}(\mu, \sigma^2)$, which implies exogeneity and homoskedasticity. Compared to time-series data where we make the normality assumption as well, but conditional on independent variables, across time. This normality assumption implies exogeneity, homoskedasticity and no serial correlation.

We need to rely in the large sample assumptions for time series compared to cross-sectional, because the small sample assumptions in TS are too restrictive and are often times too hard to fulfill. With cross-sectional data, this is not so much the case.

Part B: Write out the ACFs and PACFs for the following processes: AR(2), MA(2), AR(3), MA(3). What is the pattern that you see?

The document attached has the ACF's and PACF's calculated.

A pattern I recognized with the ACF's of an AR(q) process is that the ACF coefficients are raised to the q order of the AR(q) process. The PACF's of an AR(q) process are the coefficient of the term y_{t-q} .

A pattern I recognized with the ACF's of an MA(q) process is that for the ACF of the q term is the coefficient of that term divided by the variance of y_t . The ACF of the $q - 1$ term follows $\frac{\alpha_{q-1} + \alpha_{q-1}\alpha_q}{\text{Var}(y_t)}$. The PACF's of an MA(q) process is the first coefficient, α_1 raised to the q power.