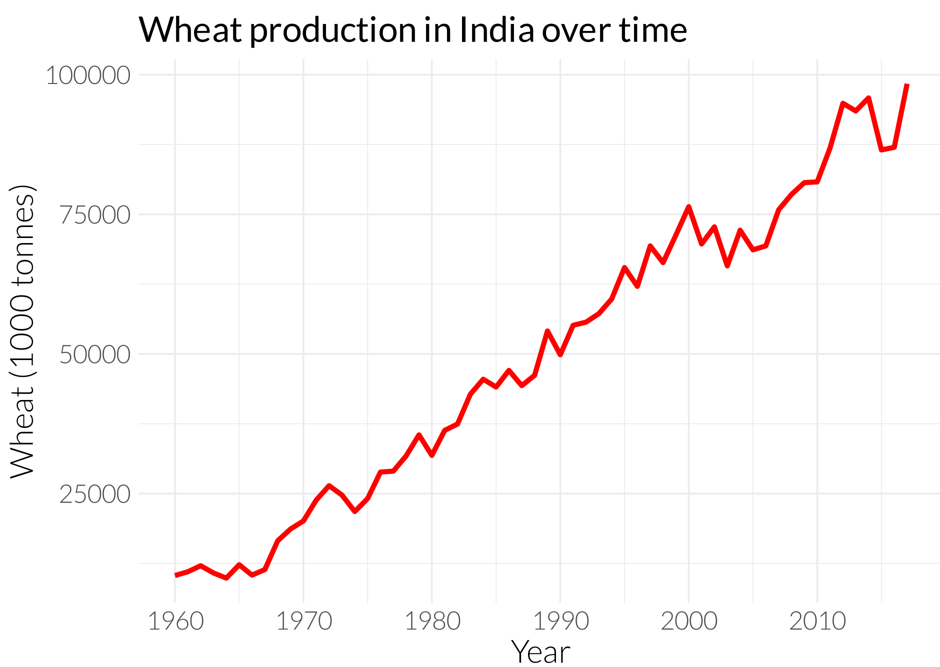
Macroeconometrics
Supervision 1

Samuel Lee

Question 1

(a)



The trend here looks approximately linear. If so, a model for the trend might be

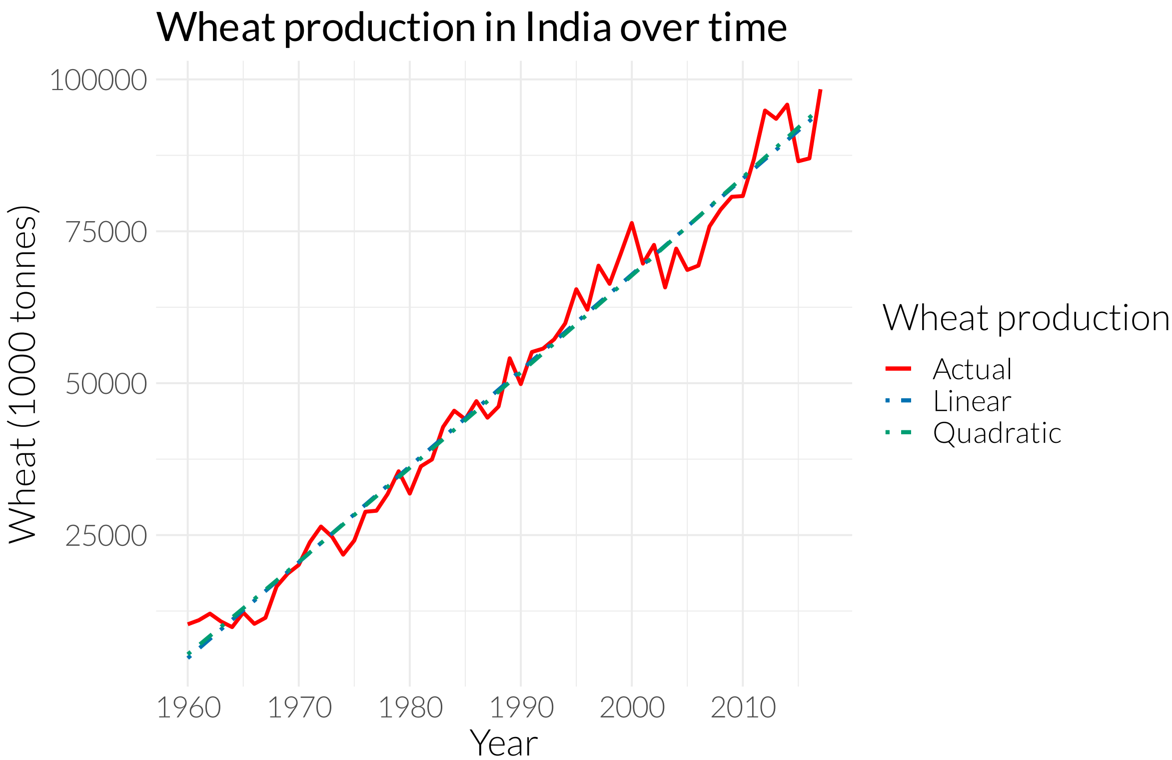
$$Wheat_t = \beta_0 + \beta_1 \, Year_t + \varepsilon_t$$

where $Year = \{0, 1, \ldots, 56, 57\} = t$ is the number of years since 1960. Thus β_0 will be the predicted wheat production at the start of the time series.

India mostly produces wheat for domestic consumption. Indian exports of wheat only took up about 500 million tonnes in 2018 out of around 1,000,000 million tonnes of production, with the bulk of exports going to the bordering countries Nepal and Afghanistan. By contrast, about 98000 million tonnes were consumed domestically (United States Department of Agriculture). If we assume that population growth in India follows an exponential path, and that wheat consumption and production is a linear function of population, we might expect an exponential trend in wheat production. However, the trend looks close to linear, which may just be a result of the time period covered; zooming in this close makes the trend look more or less linear. The plot of India's population since 1960 only looks vaguely exponential, and furthermore the plot of $\ln(Wheat_t)$ over time looks obviously non-linear compared to above. So a linear approximation may not be too far off as long as we do not extend the model too far into the past or future.

(b)

Estimating the linear and quadratic trend models, we get the following output:



The two estimated trends are so similar that it is probably difficult to tell there are two lines there when the picture is black-and-white. This, together with the previous observation that the trend looks roughly linear, probably means that adding another parameter to estimate a quadratic trend doesn't achieve much in terms of a better fit. The first instinct would be that the linear model will have a lower AIC and BIC value compared to the quadratic model; the AIC and BIC essentially gives a 'better' (lower) value for models with a better fit and a 'worse' (higher) value to models with more estimated parameters, all else equal.

As expected, the values shown in Table 1 suggest that the linear model (the model with just t as a parameter) is preferred to the quadratic model under both criteria[1]. Furthermore, we can see that the magnitude of the coefficient on $t2$ is a measly 1.194 and not statistically significant (it is even lower than the standard error of the estimate), vindicating the prior observation that the trend looks basically linear. So from our visual inspection it makes sense that a linear model would be preferred under these two criteria since the quadratic model 'wastes' one parameter on a trend that basically looks linear anyway.

(c)

The preferred model is the linear model, which as mentioned above is

$$Wheat_t = \beta_0 + \beta_1 \, Year_t + \varepsilon_t$$

[1]The AIC and BIC values will differ slightly from those calculated by Stata, but this won't affect comparisons between two different models under the same criterion. For example, Stata calculates the BIC value as 1127.541 for the linear model and 1131.232 for the quadratic model. The difference between the two is the same as in the values reported here.

Table 1: Results of estimation of linear and quadratic models

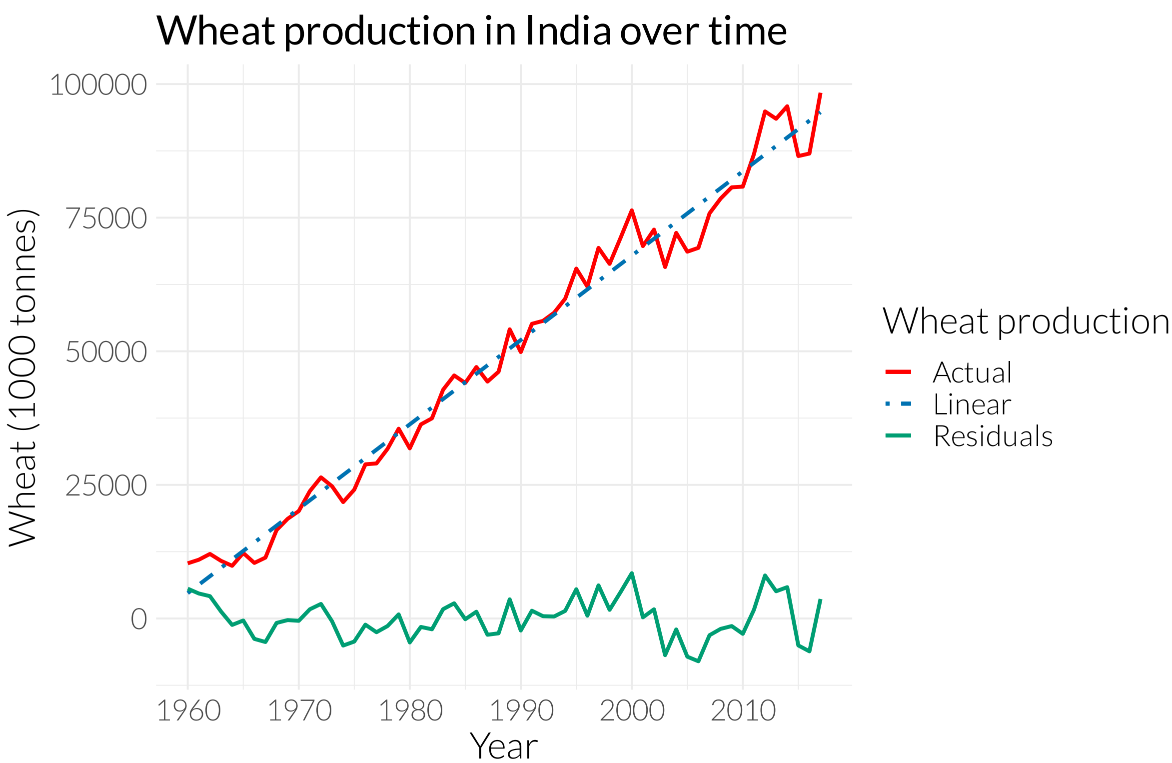
	Dependent variable:	
	wheat	
	(1)	(2)
t	1,578.695***	1,510.663***
	(29.986)	(118.761)
t2		1.194
		(2.015)
Constant	4,745.114***	5,380.077***
	(991.112)	(1,463.968)
Observations	58	58
R^2	0.980	0.980
Adjusted R^2	0.980	0.980
Akaike Inf. Crit.	1,125.420	1,127.051
Bayesian Inf. Crit.	1,131.601	1,135.293
Residual Std. Error	3,822.950 (df = 56)	3,845.305 (df = 55)
F Statistic	2,771.868*** (df = 1; 56)	1,370.042*** (df = 2; 55)
Note:		*p<0.1; **p<0.05; ***p<0.01

To come up with a meaningful forecast, we use the estimated parameters $\hat{\beta}_1 = 1578.695$ and $\hat{\beta}_0 = 4745.114$. When predicting wheat production for 2018, we are trying to predict what happens at $t = 58$. So the point estimate is $\hat{\beta}_0 + 58\hat{\beta}_1 = 4745.114 + 1578.695 \cdot 58 = 96309.424$. The 95% confidence interval for this prediction is $[96309.424 - 1.96(3822.950), 96309.424 + 1.96(3822.950)] = [88816.442, 103802.406]$, since the residual standard error is 3,822.950, but this assumes that ε_t is normally distributed with variance σ^2, and furthermore that $\text{Var}[\varepsilon_t|t] = \sigma^2\ \forall t$. Also, we are ignoring the uncertainty involved in estimating the parameters and the standard errors, although we could fix the latter by using the t-distribution instead.

(For what it's worth, the original dataset was recently updated and India's wheat production in 2018 was 99700 thousand tonnes.)

(d)

Computing the residuals as the actual volume of wheat production minus the volume of wheat production predicted by the linear trend, the plot of the residuals is as follows:



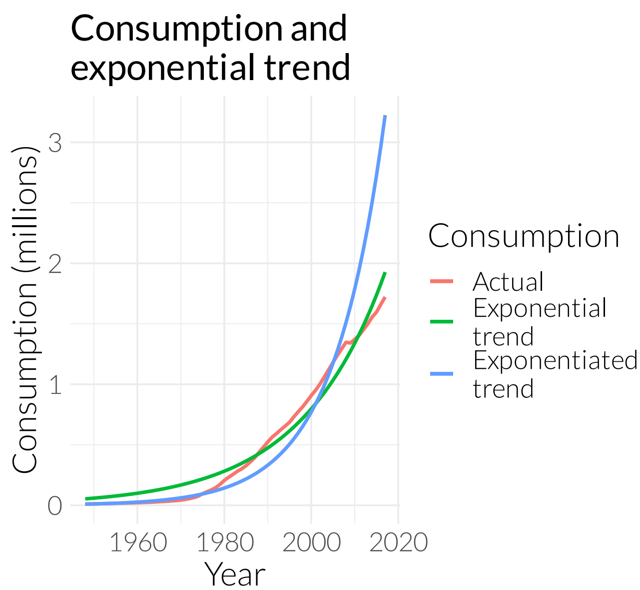
Visually it seems that there's some degree of autocorrelation in the residuals, especially after 2000. As a side note, many of the deviations below trend happened during periods where India came close to a famine due to droughts (the Bihar drought in 1967, the Maharashtra drought in 1972, the West Bengal drought in 1979, and the other Maharashtra drought in 2013). Somewhere around 2000 the Indian government kept the Minimum Support Price (effectively a price floor) relatively low which dampened wheat production as reflected above.

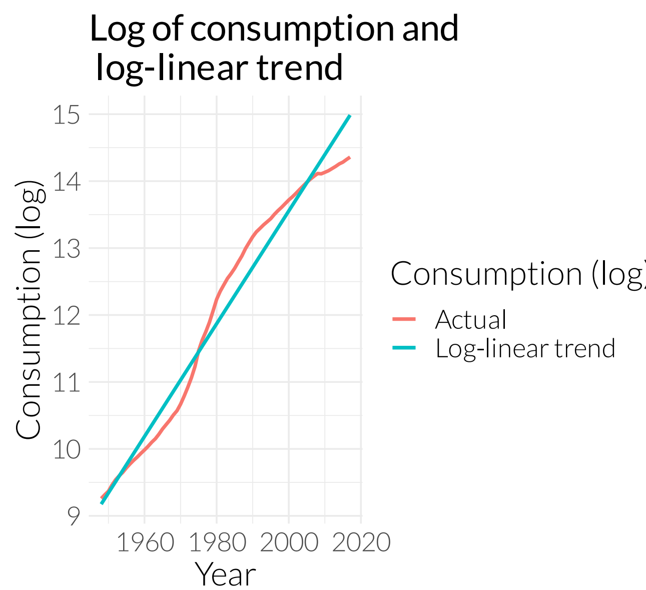
The Durbin-Watson (DW) statistic tests for serial correlation in ε_t over time. The DW statistic is $\frac{\sum_{t=2}^{T}(e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2}$, and turns out to be approximately 1.093436. The null hypothesis is that there is no serial correlation between ε_t and ε_{t-1}, and is rejected if the DW statistic is "much smaller than" 2. According to the DW significance tables, with 1 regressor and around 60 observations, the lower bound for the 1% critical value is 1.382. The DW statistic here is smaller, so we reject the null hypothesis that there is no serial correlation in the regression disturbances, as conjectured above. Being below the critical range suggests there may be positive serial correlation, and the economic reasons include some of the points raised above; the serial correlation may reflect persistence of weather conditions or government policies.

Question 2

(a)

The model for the nonlinear trend is $T_t = \beta_0 e^{\beta_1 \cdot t}$ and the model for the log-linear trend is $\ln T_t = \ln \beta_0 + \beta_1 t$. In the non-linear trend, β_0 is estimated as around 5.353×10^4 and β_1 is estimated as around 5.194×10^{-2}. In the log-linear trend, the intercept is estimated as around 9.09091 and the coefficient on time is estimated as around 0.08422. Exponentiating this result, we get the following plot (with the log-linear trend plotted for reference):





The obvious difference is that the exponentiated log-linear trend consistently underestimates consumption from the mid-1970s to about 2005, and grossly overestimates consumption from 2005 onwards. Exponentiating the predicted value for the log of consumption leads to a biased estimate, since

$$
\begin{aligned}
\mathrm{E}[y|x] &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{u^2}{2\sigma^2}} e^{\beta_0+\beta_1 x_1+\ldots+\beta_k x_k+u} \mathrm{d}x \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{u^2}{2\sigma^2}} e^{x^\intercal \beta+u} \mathrm{d}x \\
&= e^{x^\intercal \beta} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{u^2-2\sigma^2 u}{2\sigma^2}} \mathrm{d}x \\
&= e^{x^\intercal \beta} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{u^2-2\sigma^2 u+\sigma^4-\sigma^4}{2\sigma^2}} \mathrm{d}x \\
&= e^{x^\intercal \beta} e^{\frac{\sigma^2}{2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(u-\sigma^2)^2}{2\sigma^2}} \mathrm{d}x = e^{x^\intercal \beta} e^{\frac{\sigma^2}{2}}
\end{aligned}
$$

More generally, because $f(x) = e^x$ is a convex function, $\mathrm{E}\left[f(x)\right] \geq f\left(E[x]\right)$ due to Jensen's inequality. Therefore the exponentiated trend is biased downwards.

Between this approach and the nonlinear least-squares method, the latter is preferred for this reason. Still, in this particular instance an even better alternative could just be the log-linear model itself. As shown above, the cyclical movements in consumption are a lot more obvious in the log-linear plot, and the parameters have a (probably) more useful economic interpretation: the coefficient on time shows the approximate percentage increase in consumption each year. It also has the advantage of having a closed-form solution, whereas the nonlinear least squares method relies on numerical optimization through an iterative process, and the estimated parameters will differ based on the different starting points one chooses for the parameters. An estimate may not even be found if one chooses a starting point that is too far from the least squares estimate, so some knowledge of the plausible range of values the parameters take is needed, which somewhat defeats the point of the exercise.

(b)

Intuitively, the AIC value is related to the mean squared error of the estimated dependent variables. In this case the dependent variables are different between the two methods; one measures consumption and the other measures the log of consumption. A mean squared error of 1 has a very different meaning when referring to consumption or the log of consumption, and so the AIC values cannot be compared.

(c)

We have that

$$MSE = \mathrm{E}[e^2] = \mathrm{E}[e^2] - \mathrm{E}[e]^2 + \mathrm{E}[e]^2 = \mathrm{Var}[e] + \mathrm{E}[e]^2$$

which is what we have to show.

An unbiased forecast is desirable, all else being equal, since it provides more information about the possible future state than if the forecast were biased and the magnitude of the bias was unknown. However, in some cases it might be worthwhile to trade off a little bit of bias for a higher efficiency. Some estimators have a lower $\mathrm{E}[e]^2$ at the cost of a higher $\mathrm{Var}[e]$ and vice versa. It may be more important in certain contexts to have a precise estimate, especially if the bias is known to be in a direction that doesn't invalidate what the forecast is being used for or if the estimator is biased but consistent. For example, if a government wants a forecast of a bad weather event in the future so as to prepare the emergency services, it will probably want a more precise forecast and will not mind too much if the forecast is slightly biased to be over-cautious.

(d)

As mentioned in (a), the latter is not an unbiased forecast. If $(\log y)_{T+h,T}$ is an unbiased forecast, this means

$$(\log y)_{T+h} = (\log y)_{T+h,T} - u$$

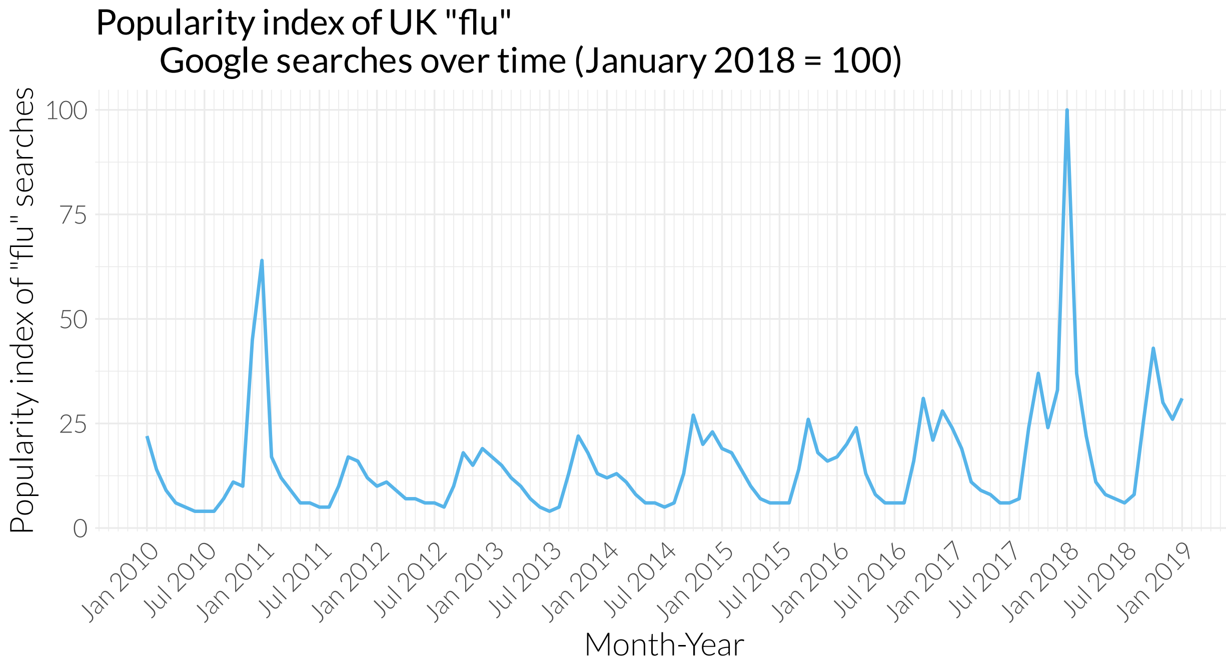
where $\mathrm{E}[u] = 0$. But the exponent is not unbiased:

$$e^{(\log y)_{T+h,T}} = e^{(\log y)_{T+h}+u} = e^u e^{(\log y)_{T+h}} = e^u y_{T+h}$$
$$\mathrm{E}[e^u y_{T+h}] \neq \mathrm{E}[y_{T+h}]$$

If u were normally distributed then we would somehow get the same result mentioned in (a), but even if it were not the left-hand side would not in general be equal to the right-hand side except in special cases (like if u took the value 0 with probability 1).

Question 3

(a)



There seems to be a seasonal pattern where the frequency of “flu” Google searches tends to peak at October. 6 out of 9 of the yearly peaks in this series were attained in October. In the years where they didn’t (2010, 2012 and 2017), the peak was obtained December or January, although there was always a local peak in October in all the years.

The exceptionally high values for December 2010, January 2011, and January 2018 are most likely due to one-off salient events: the return of “swine flu” (H1N1) in late 2010 just one year after the 2009 pandemic, and the so-called “Aussie” flu (H3N2) in early 2018 (although the strain has been prevalent for many years before that; it is only the media coverage and the term “Aussie flu” that became widespread in 2018, and only in the UK).

(b)

We normalize January 2010 to $t = 0$. With a linear trend, a seasonal dummy variable (1 for each month), and dummies for December 2010, January 2011, and January 2018, the model is

$$y_t = \beta t + \sum_{i=1}^{12} \gamma_i M_{it} + \delta_1 D2010_t + \delta_2 J2011_t + \delta_3 J2018_t + \varepsilon_t$$

where for example $M_{1t} = 1$ if $t = 0, 12, 24, \dots$ and $M_{1t} = 0$ otherwise. Estimating the model yielded the results in Table 2.

From Table 2, it seems that after partialling out the disproportionate effects in December 2010, January 2011, and January 2018, the month which is most highly associated with increased “flu” search queries is October, with the largest estimate M10 of 19.311. If Google queries about the flu are well-correlated with actual incidences of flu (which may be a questionable assumption given that these queries shot through the roof in January 2018 mostly due to media attention), then it would appear that October was the month where one would be most likely to contract the flu. The “least dangerous” month based on these estimates would be July.

Table 2: Estimated seasonal and trend components

	Dependent variable:
	flu
t	0.113*** (0.013)
M1	12.893*** (1.628)
M2	12.681*** (1.524)
M3	8.123*** (1.530)
M4	3.454** (1.535)
M5	1.008 (1.541)
M6	−0.216 (1.547)
M7	−0.774 (1.553)
M8	−0.442 (1.559)
M9	8.444*** (1.565)
M10	19.331*** (1.571)
M11	12.552*** (1.577)
M12	13.899*** (1.695)
D2010	29.857*** (4.457)
J2011	49.750*** (4.435)
J2018	76.250*** (4.435)
Observations	109
R^2	0.963
Adjusted R^2	0.957
Residual Std. Error	4.149 (df = 93)
F Statistic	152.095*** (df = 16; 93)
Note:	*p<0.1; **p<0.05; ***p<0.01

(c)

The Durbin-Watson statistic is calculated as 0.95957. With 16 regressors and around 100 observations, the statistic is below the 1% minimal bound for positive serial correlation which is 1.185, meaning the null hypothesis of zero serial correlation will also be rejected at the 5% level. This means it seems unlikely that the residuals are independent and random, and there is possibly some cyclical component in the process even after removing the trend and seasonal components.

Question 4

(a)

Covariance stationarity requires that

$$\mathrm{E}[y_t] = \mu,\ \mathrm{Cov}[y_t, y_{t-\tau}] = \gamma(\tau),\ \gamma(0) = \mathrm{Var}(y_t) < \infty\ \forall t$$

For odd t, the expected value is 0, and for even t, the expected value is also 0. This satisfies the first condition: $\mathrm{E}[y_t] = \mu = 0\ \forall t$. Since $\mu = 0$, the autocovariance $\mathrm{Cov}(y_t, y_{t-\tau})$ is equal to $\mathrm{E}[y_t y_{t-\tau}]$. If $\tau \neq 0$, then y_t and $y_{t-\tau}$ are independent and the expectation of the product is the product of the expectation, which is 0. The variance of y_t is

$$\mathrm{E}[y_t^2] - \mathrm{E}[y_t]^2 = \mathrm{E}[y_t^2] = \begin{cases} \frac{1}{4} \cdot 3^2 + \frac{3}{4}(-1)^2 = 3 & \text{if } t \text{ is odd} \\ \frac{3}{4} + \frac{1}{4} \cdot (-3)^2 = 3 & \text{if } t \text{ is even} \end{cases} = 3\ \forall t$$

which means the other two conditions are satisfied. Therefore the process is covariance stationary.

To see if $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} y_t$ converges, we calculate

$$\mathrm{E}[|y_t^3|] = \begin{cases} \frac{1}{4} \cdot 3^3 + \frac{3}{4} = \frac{15}{2} & \text{if } t \text{ is odd} \\ \frac{3}{4} + \frac{1}{4} \cdot 3^3 = \frac{15}{2} & \text{if } t \text{ is even} \end{cases} = \frac{15}{2}\ \forall t$$

As given, y_t is independent but not identically distributed, and a sufficient condition for the central limit theorem to apply is given by the Lyapunov condition

$$\lim_{T\to\infty} \frac{1}{\left(s_T^2\right)^{\frac{3}{2}}} \sum_{t=1}^{T} \mathrm{E}[|y_t|^3] = 0$$

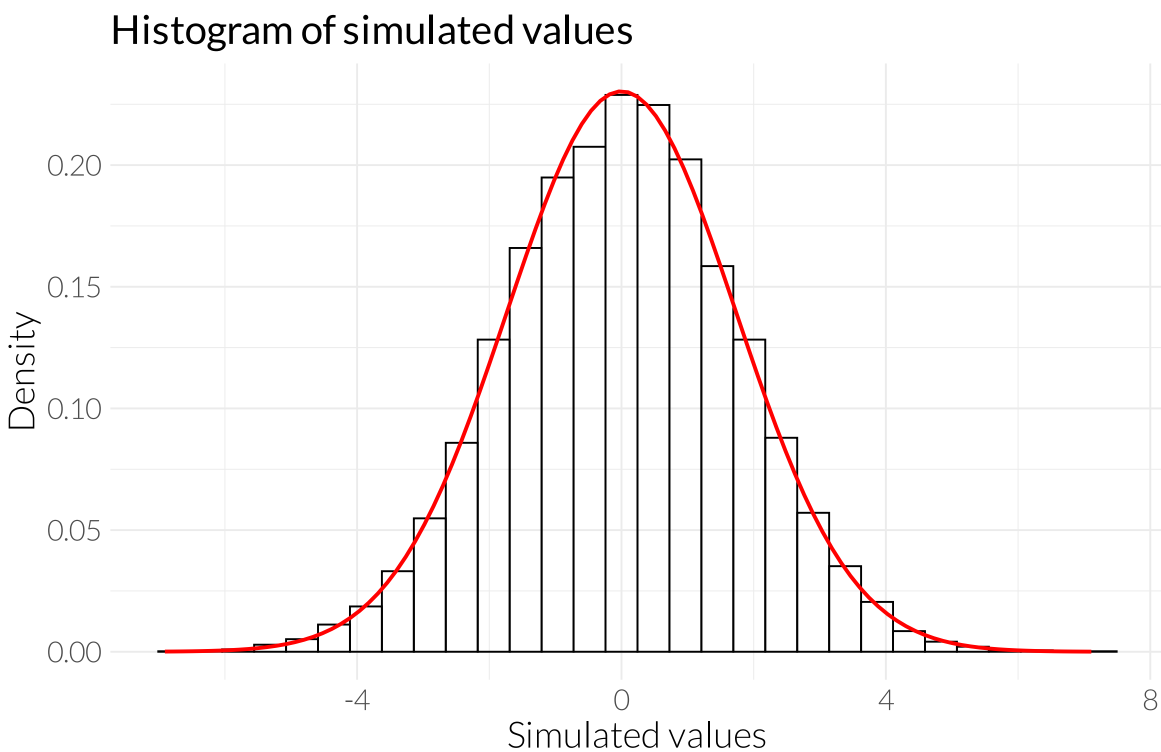
where $s_T^2 = \sum_{i=1}^{T} \mathrm{E}[y_t^2]$. In this context the above condition is satisfied:

$$\lim_{T\to\infty} \frac{1}{\left(s_T^2\right)^{\frac{3}{2}}} \sum_{t=1}^{T} \mathrm{E}[|y_t|^3] = \lim_{T\to\infty} \frac{1}{(3T)^{\frac{3}{2}}} \cdot \frac{15T}{2} = \lim_{T\to\infty} \frac{5}{2\sqrt{3T}} = 0$$

And therefore, by the Lyapunov central limit theorem,

$$\frac{1}{\sqrt{\sum_{t=1}^{T} \mathrm{E}[y_t^2]}} \sum_{t=1}^{T} y_t = \frac{1}{\sqrt{3T}} \sum_{t=1}^{T} y_t = \frac{1}{\sqrt{3}} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} y_t \xrightarrow{D} N(0,1)$$

which implies $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \xrightarrow{D} N(0,3)$. If the algebra is unconvincing, below is the density histogram of simulated values where T is 1 million and $\frac{1}{T} \sum_{t=1}^{T} y_t$ is sampled ten thousand times, with the density function of $N(0,3)$ overlaid.



(b)

The values taken on by x_t are now deterministic given t; they take on the realized (known) values of y_1 and y_2. The expected value is no longer 0 at all t (or for any t for that matter). $\mathrm{E}[x_t] = y_1$ for odd t and $\mathrm{E}[x_t] = y_2$ for even t, and there is no way for them to be equal. So $\mathrm{E}[x_t]$ now does depend on t and the process is no longer weakly stationary.

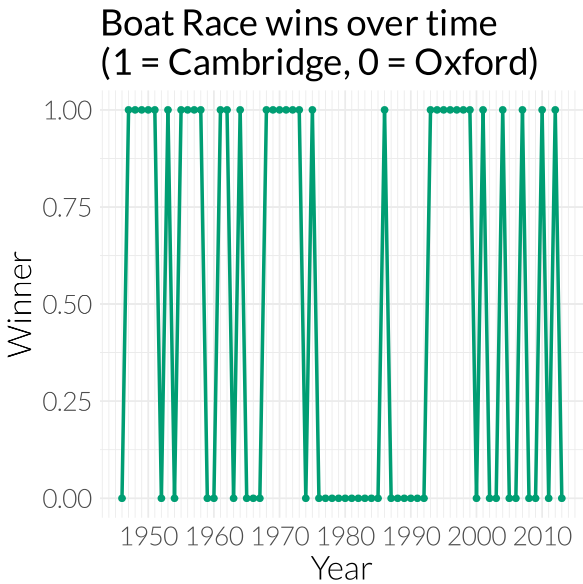
(c)

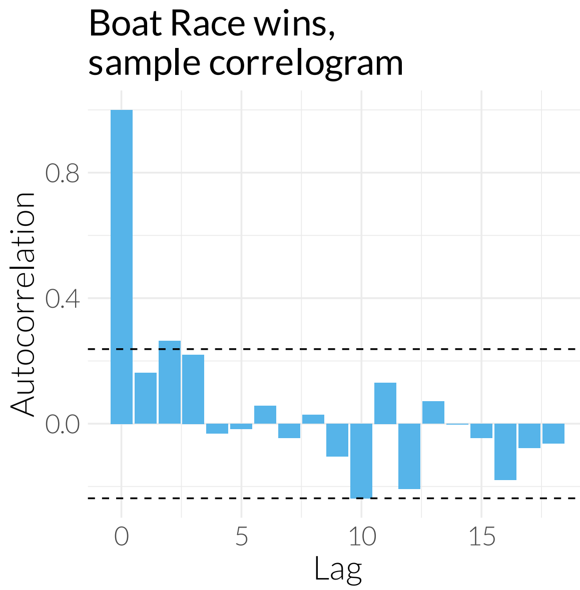
As mentioned before, x_t is now deterministic, and $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} x_t$ has no distribution other than at the value it takes at each T. Furthermore the process is not covariance stationary and the expected value μ for each x_t depends on whether t is odd or even, so the CLT would not apply.

Question 5

(a)

The time series plot together with the sample correlogram of $Winner_t$ are as follows





The sample correlogram only has $\hat{\rho}(0) = 1$ being very obviously beyond the confidence intervals within which we expect 95% of the sample correlations to fall if the data were white noise. Still, there may be some cause for concern particularly at around $\hat{\rho}(1)$, $\hat{\rho}(2)$, and $\hat{\rho}(3)$ where the sample autocorrelation is persistently positive (with $\hat{\rho}(2)$ being slightly above the confidence interval band). It is plausible that Boat Race wins are slightly correlated with a lag of around 3 years before the correlation tapers off rapidly; a win in any year may indicate a stronger-than-usual team (or weaker-than-usual opponent), and after about 3 years most of that original winning (or losing) team would have graduated.

(b)

The Ljung-Box and Box-Pierce tests both test whether groups of autocorrelations are jointly significant. The theory goes that when a process is white noise, $\hat{\rho}(\tau)$ are asymptotically independent and normally distributed with mean 0 and variance $\frac{1}{T}$. This means

$$\hat{\rho}(\tau) \sim N\left(0, \frac{1}{T}\right)$$

$$\sqrt{T}\hat{\rho}(\tau) \sim N(0,1)$$

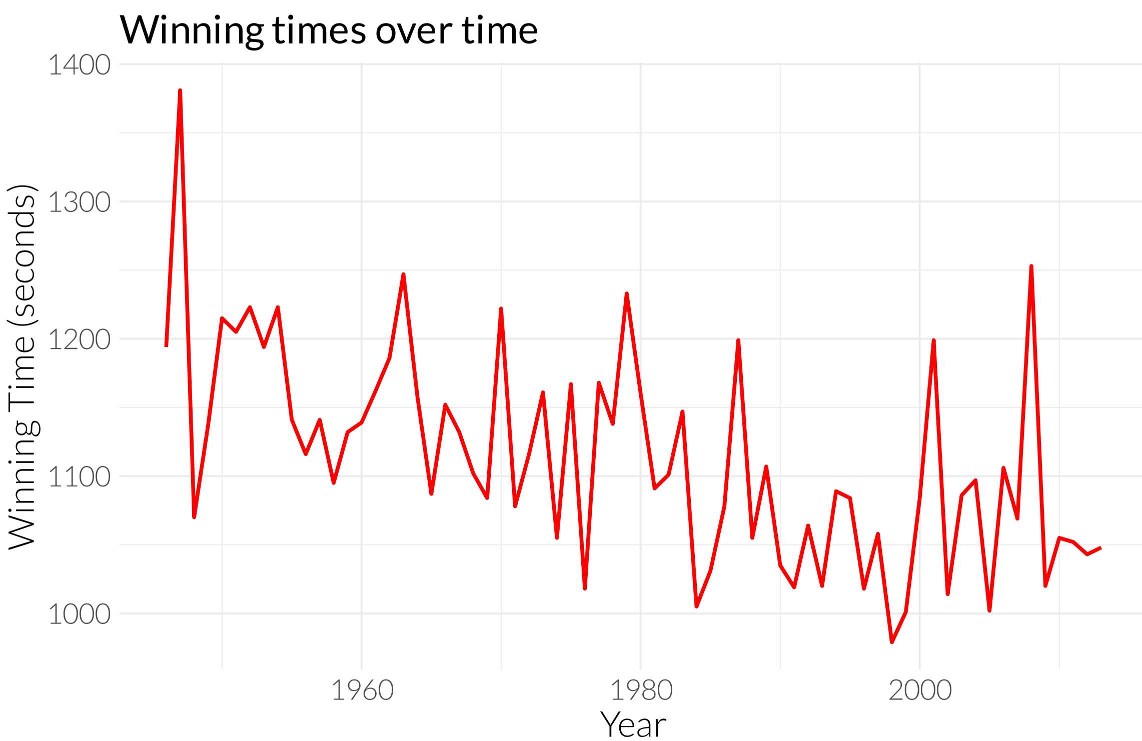
$$T\hat{\rho}(\tau)^2 \sim \chi^2(1)$$

At different τ, the χ^2 distributeions are asymptotically independent, and the sum of m $\chi^2(1)$ variables has a $\chi^2(m)$ distribution. The Ljung-Box makes a slight modification to this with the test statistic $Q_{LB} = T(T+2)\sum_{\tau=1}^{m} \frac{1}{T-\tau}\hat{\rho}(\tau)^2$.

Using the computer to do all this for us, we get $Q_{LB} = 10.503$. From the statistical tables, a test-statistic with a $\chi^2(4)$ distribution that is slightly more than 10.5 has a p-value of at most 0.0328. This means the null hypothesis of the joint insignificance of $\hat{\rho}(\tau)$ for τ up to 4 is rejected; the results are suggestive that there may be some autocorrelation in the process. We reject the hypothesis that the Winner series is white noise.

(c)

The time plot of $WinTime_t$ is as follows



There is an obvious declining trend in the winning times, although there may be a hint of a slightly increasing trend from about 2000.

(d)

With $WinTime_t = \gamma + \delta \times Year_t + v_t$, we have that

$$DWT_t = WinTime_t - WinTime_{t-1} = \delta(Year_t - Year_{t-1}) + (v_t - v_{t-1})$$

In which case the fixed effect γ is differenced out. The autocovariance of DWT_t with lag j is

$$\begin{aligned}&\text{Cov}[DWT_t, DWT_{t-j}]\\ =&\text{Cov}[\delta(Year_t - Year_{t-1}) + (v_t - v_{t-1}), \delta(Year_{t-j} - Year_{t-j-1}) + (v_{t-j} - v_{t-j-1})]\end{aligned}$$

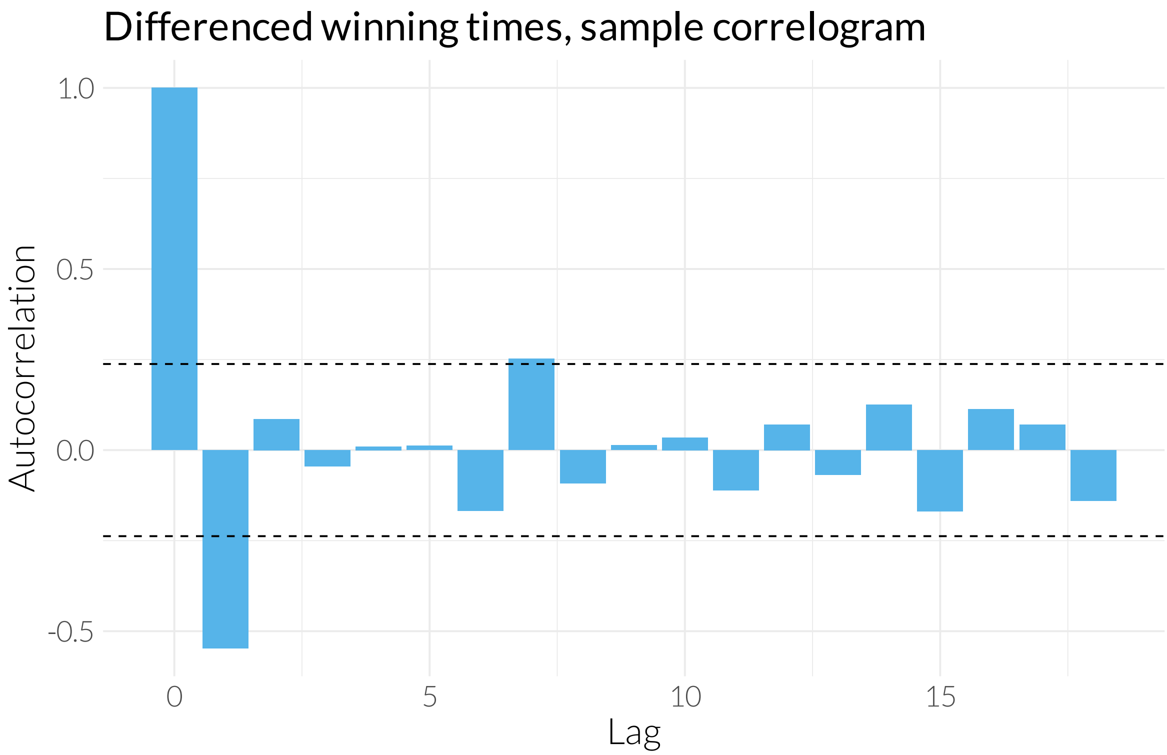
$Year_t$ is a non-stochastic variable, so its variance and covariance with any other variable is 0. Therefore,

$$\begin{aligned}\text{Cov}[DWT_t, DWT_{t-j}] &= \text{Cov}[v_t - v_{t-1}, v_{t-j} - v_{t-j-1}]\\ &= \text{Cov}[v_t, v_{t-j}] - \text{Cov}[v_t, v_{t-j-1}] - \text{Cov}[v_{t-1}, v_{t-j}] + \text{Cov}[v_{t-1}, v_{t-j-1}]\end{aligned}$$

If $j = 0$, then the middle two terms are zero and the expression is equal to $2\sigma_v^2$. If $j = 1$, then all but the third term are zero, and $\text{Cov}[DWT_t, DWT_{t-j}] = -\sigma_v^2$. For any $j > 1$, all the terms are 0. The autocorrelation function is just the autocovariance function at lag τ divided by the autocovariance function at lag 0 (which is the variance). Therefore, $\rho_j = \begin{cases} -\frac{1}{2} & \text{if } j = 1 \\ 0 & \text{if } j > 1 \end{cases}$. This means that the differenced winning times would be negatively correlated with their one-period lagged values. This makes sense, since a higher-than-usual v_t will make a positive DWT_t and a negative DWT_{t+1} more likely.

(e)

The sample correlogram is as follows



As predicted, the sample autocorrelation at lag 1 is negative, and even better, the magnitude is very close to what was expected $\left(-\frac{1}{2}\right)$. The autocorrelation at longer lag periods are somewhat within the 95% confidence interval bands.