Time Series Analysis and Forecasting (ECOM30004): Assignment 1

Will Mackey
790114

2018
August

1 Conceptual question

1.1

The first structural-break model is:

$$y_t = \alpha_0 + \alpha_1 DI_t + U_{1t} \qquad \text{where } U_{1t} \text{ is noise; } \textit{(i)}$$

$$DI_t = \begin{cases} 0 & \text{if} \quad t \leq T_B \\ 1 & \text{if} \quad t > T_B \end{cases} \qquad \text{where } 1 < T_B < T$$

When the structural break occurs for model (i), the change in expected output is the post-break expectation $E[y_{t,DI_t=1}]$ less the pre-break expectation $E[y_{t,DI_t=0}]$, shown below.

From:

$$\begin{aligned} E[y_{t,DI_t=0}] &= E[\alpha_0] + E[U_{1t}] && \text{where } E[U_{1t}] = 0 \\ &= \alpha_0 && \text{as } \alpha_0 \text{ is a constant} \end{aligned}$$

to:

$$\begin{aligned} E[y_{t,DI_t=1}] &= E[\alpha_0 + \alpha_1 + U_{1t}] \\ &= E[\alpha_0] + E[\alpha_1] + E[U_{1t}] && \text{where } E[U_{1t}] = 0 \\ &= \alpha_0 + \alpha_1 \end{aligned}$$

giving:

$$E[y_{t,DI_t=1}] - E[y_{t,DI_t=0}] = \alpha_1$$

Or, the difference in expected y_t is α_1. This shows that model (i), which is simulated in Figure 1 alongside model (ii), could represent a time-invariant variable that changes *levels* due to a shock. For example, productivity growth (y_t) of an established industry that experiences a permanent negative shock $(a_1 < 0)$ when a competitor enters the market at time T_B.

The second structual-break model is:

$$y_t = \beta_0 + \beta_1 t + \beta_2 DI_t + U_{2t} \qquad \text{where } U_{2t} \text{ is noise, } \textit{(ii)}$$

$$DI_t = \begin{cases} 0 & \text{if} \quad t \leq T_B \\ t - T_B & \text{if} \quad t > T_B \end{cases} \qquad \text{where } 1 < T_B < T$$

Similarly, when the structual break occurs, the expected output changes by the post-break expectation $E[y_{t,DT_t=t-T_B}]$ less the pre-break expectation $E[y_{t,DT_t=0}]$.

From:

$$
\begin{aligned}
E[y_{t,DT_t=0}] &= E[\beta_0 + \beta_1 t + U_{2t}] && \text{where } E[U_{2t}] = 0 \\
&= \beta_0 + \beta_1 t
\end{aligned}
$$

to:

$$
\begin{aligned}
E[y_{t,DT_t=0}] &= E[\beta_0 + \beta_1 t + \beta_2(t - T_B) + U_{2t}] && \text{where } E[U_{2t}] = 0 \\
&= \beta_0 + \beta_1 t + \beta_2(t - T_B)
\end{aligned}
$$

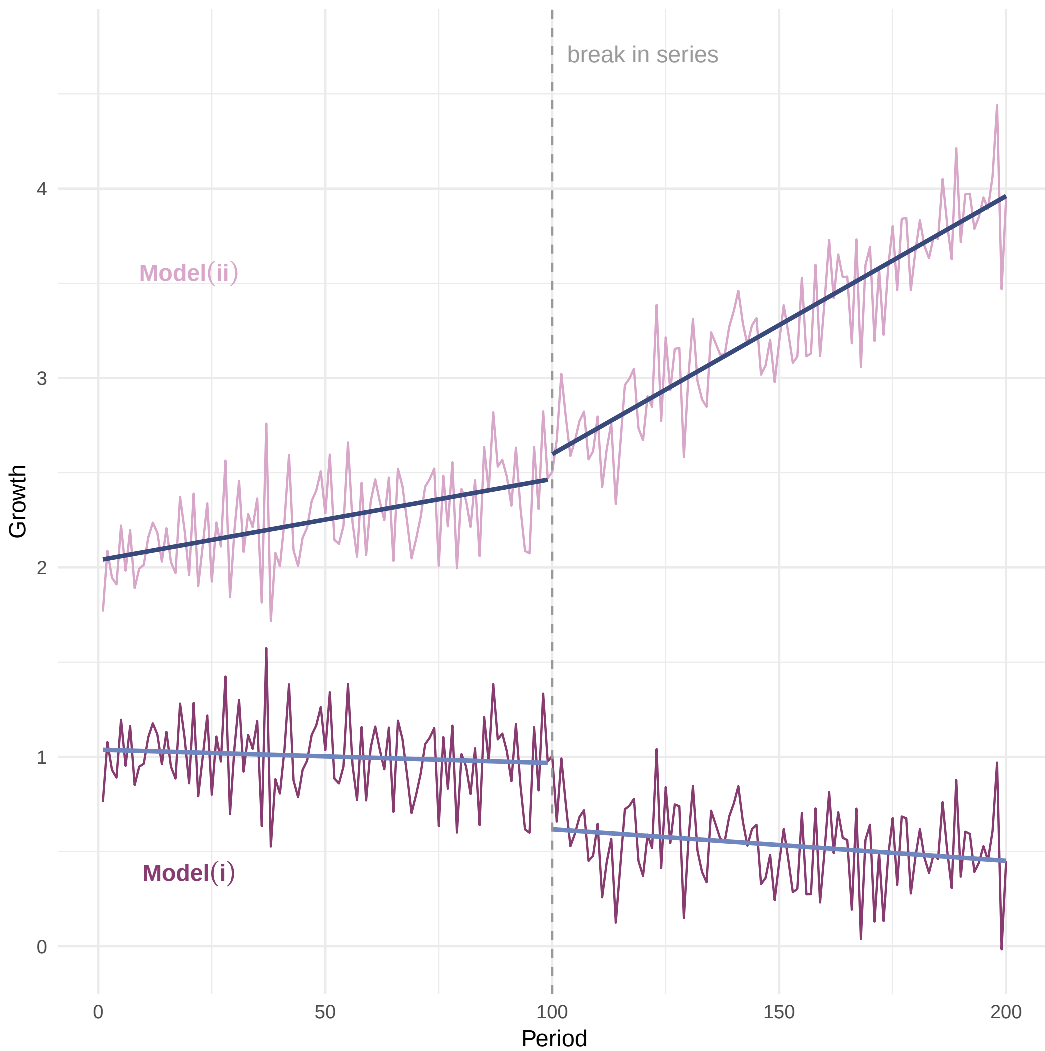
giving the change:

$$
E[y_{t,DT_t=(t-T_B)}] - E(y_{t,DI_t=0}) = \beta_2(t - T_B)
$$

Or, the difference in expected y_t is the estimated time-dependant linear trend with coefficent β_2. Unlike model (i), model (ii) is time-variant with a time-dependant structural break: the pre-break trend is dependant on time, and this 'relationship' with time changes significantly after the break. It could represent high- and increasing-growth ($\beta_1 > 0$) in university student numbers at Australian institutions after a the introduction of the demand-driven system at time T_B. This is, very roughly, shown in Figure 1.[1]

[1] One would have to assume monthly data-collection on enrolments and a lack of a semester-based enrolment system for this figure to match the example well. Strong imagination required.

Figure 1: Models (i) and (ii)



Note: The two models simulated with the same normally distributed errors over 200 time periods. The structural break occurs at t=100. Parameters for model (i): $\alpha_0 = 1$ *and* $\alpha_1 = -0.5$*; for model (ii):* $\beta_0 = 2$, $\beta_1 = 0.005$ *and* $\beta_2 = 0.01$*. Linear trends with breaks are shown for model (i) and model (ii).*

1.2

Model A is time-variant with one structural break and can be expressed as

$$p_t = \alpha_0 + \alpha_1 t + \alpha_2 T_{B,1} + U_{1t}$$

Model B *likely* has six structural breaks, denoted in the following general-form model as $T_{B,1}$ through $T_{B,6}$.[2]

$$\begin{aligned} p_t = \beta_0 + \beta_1 t + \beta_2 T_{B,1} \\ + \beta_3 T_{B,2} \\ + \beta_4 T_{B,3} \\ + \beta_5 T_{B,4} \\ + \beta_6 T_{B,5} \\ + \beta_7 T_{B,6} \\ + U_{2t} \end{aligned}$$

Model B increases the information in (number of variables of) the model, and *necessarily* increases its "fit" (R^2).[3] But this is at the cost of degrees of freedom, and will produce an overfitted model. The coefficients of $T_{B,n}$ would the noise of the data rather than any proper relationship with population events. As the number of breaks approaches the number of observations $n \to N$ we lose the ability to model structural change; it would be like introducing a dummy variable for each period of observation (day, in this case), which would provide a perfect summary of the Yahoo share-price in the past but provide no information of any other trend. The model would fail to be deterministic — have a clear relationship with time t — as it is crowded-out by the every-day-break variables.

[2] Although it may have seven if there is one at M12; but we'll assume six for the remainder of the question.

[3] See *Econometrics 1* for a detailed discussion.

2 Empirical question

2.1

Figure 2: Daily log-volume of shares traded on the S&P/ASX 200

Note: The linear trend without a break is shown in the first panel in blue. The daily log-volume is repeated in the second panel, with a pre- and post-Chi-X trend in dark blue

The variable volume $volume_t$ is units of thousands of shares traded daily. The log of volume $log(volume_t)$ is plotted against time in Figure 2. The time-series show a deterministic trend of log-volume traded. there is are clear relationships with time. There is strong growth in trading volume between mid-2000 and late-2009. From 2010 there is a 'tempering' for 18 months followed by a decrease in trading between mid-2011 and late-2016, the end of the series.[4] From the plot, there is a clear relationship between time and log-volume traded.

[4]This decrease is notable given the population increase in Australia — a positive contributing factor of trading-volume — over this period. It might be interesting to look at, say, the trend in trading-volume per person.

2.2

A model for log volume is shown in (1) below.

$$
\begin{aligned}
\log vol_t = \beta_0 + \beta_1 t + \beta_2 DT_t &+ \beta_{3,1} Mon_t \\
&+ \beta_{3,2} Tue_t \\
&+ \beta_{3,3} Wed_t \\
&+ \beta_{3,4} Thu_t \\
&+ \beta_{3,5} Fri_t \\
&+ U_t
\end{aligned} \tag{1}
$$

Model (1) allows for three broad effects: the overall deterministic structure is modelled through t; DT_t allows for a potential structural break, visually shown in the second panel of Part 2.1, after the introduction of the Chi-X trading platform; and day-of-the-week effects are modelled through the dummy variables *Mon*-*Fri*, where *Mon* takes a value of 1 if the day is is Monday, *Tue* for Tuesday, and so on.[5] U_t is white noise.

[5] Monday is the day-of-the-week chosen to be excluded due to autocorrelation. Note also that seasonal effects would add to the model, but are excluded for simplicity.

2.3

Model was estimated in `R`, and its results are presented in Table 1.[6]

Table 1: Results of Model (1)

	Dependent variable:
	ln(*vol*)
	(1)
t	0.00046***
	(0.00001)
dt	-0.00075***
	(0.00001)
tue	0.1639***
	(0.01435)
wed	0.20808***
	(0.01434)
thu	0.2514***
	(0.01435)
fri	0.21555***
	(0.01442)
Constant	12.31408***
	(0.01386)
Observations	4,123
R^2	0.639827
Adjusted R^2	0.639302
F Statistic (df = 6; 4116)	1218.64***
Note:	*p<0.1; **p<0.05; ***p<0.01

Monday blues

Monday (Mon_t) is evidenty a slow day for trading: at the 5 per cent confidence level there is significantly more trading on all other weekdays. The model estimates that average trading volume on Mondays is $e^{12.3141} = 222810.27$ units traded. Relative to Monday, the trading volume for other weekdays is:

- Tuesday: $(e^{0.1639} - 1) \times 100\% = 17.81\%$ higher.
- Wednesday: $(e^{0.2081} - 1) \times 100\% = 23.13\%$ higher.
- Thursday: $(e^{0.2514} - 1) \times 100\% = 28.58\%$ higher.
- Friday: $(e^{0.2156} - 1) \times 100\% = 24.05\%$ higher.

The Chi-X drag

The volume-traded increased strongly between mid-2000 and 2011 (see Figure 2), and this analysis agrees: before Chi-X was introduced the average growth to trading was $(e^{0.0005} - 1) \times 100\% = 0.046\%$ per trading-day. The introduction of Chi-X reversed this growth. Post-Chi-X, the trend in average growth dropped to $(e^{(0.0005-0.0007)} - 1) \times 100\% = -0.028\%$ per trading-day.[7]

[6] Full, reproducable code for this assignment can be found at `www.github.com/wfmackey/timeseries-coursework/tree/master/assignment1` (only made public after the due date).

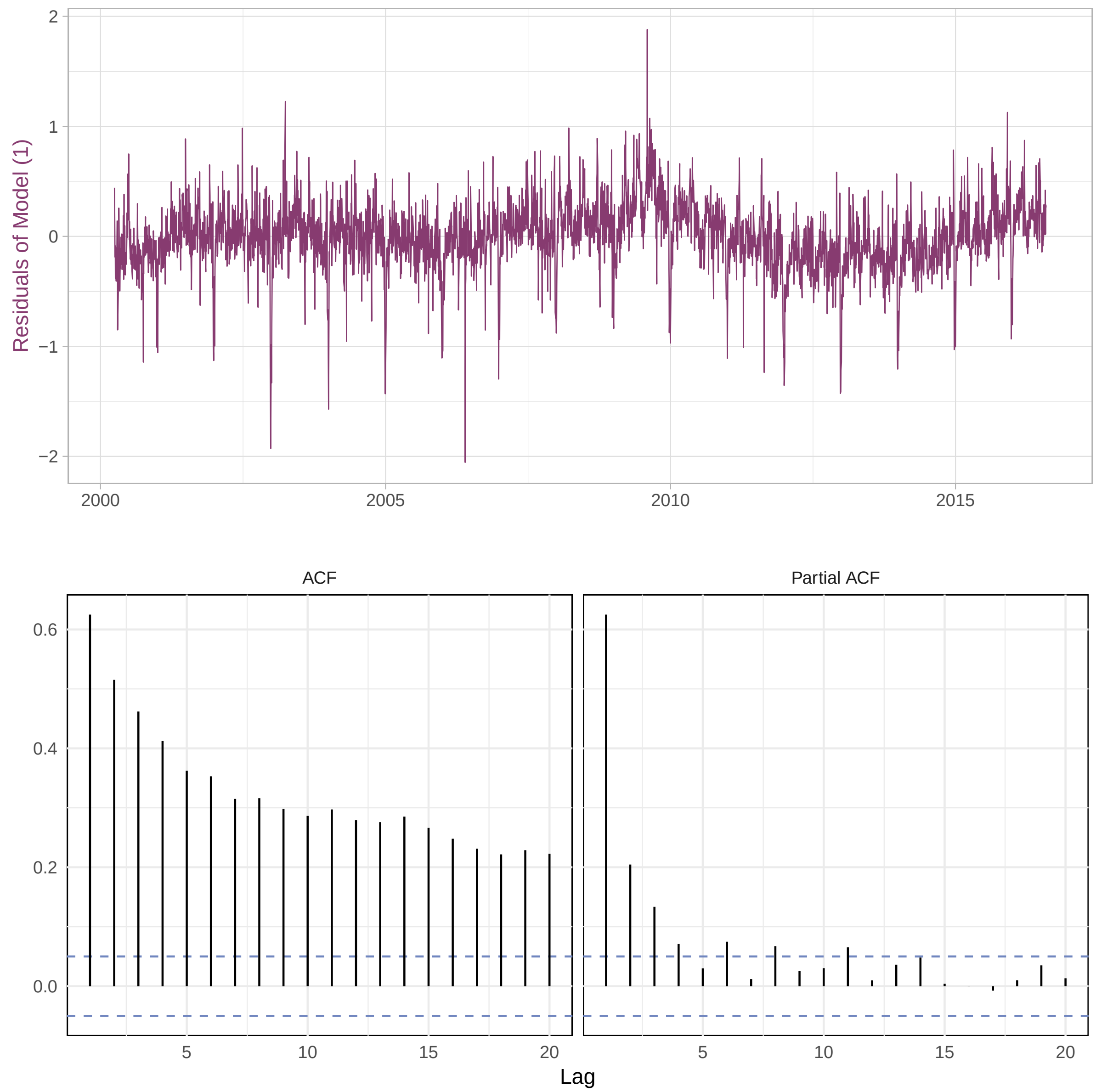
[7] This is equal to an annualised increase of 18.4% before Chi-X, and -9.8% after.

2.4

The residuals, alonside the ACF and partial ACF (PACF) of the residuals, for model (1), are plotted in Figure 3. The ACF and PACF are plotted to 20 lags. The ACF decays *reasonably* quickly to the 10^{th} lag before tapering off, indicating that there remain frequent deviations from the trend. This can be an indication of stationarity.[8]

The PACF indicates there are significant direct effects up to the fourth lag, with further effects at lags 6, 8 and 11. There is clear evidence of autocorrelation. The first lag effect for the ACF and PACF[9] is small – 0.625 – which is an indicator of stationarity.

Figure 3: Residuals and ACF/Partial ACF for model (1)



[8]But this indication will be ignored for this paper.
[9]The first lags for ACF and PACF are necessarily the same

2.5

Newey-West standard errors can be used to conduct inference in the presence of autocorrelation.[10] Newey-West standard errors for model (1) are presented in Table 2, alongside the uncorrected standard errors.

Table 2: Results of Model (1) with uncorrected and Newey-West standard errors

	Dependent variable:	
	$ln(vol)$	
	Uncorrected standard errors	Newey-West standard errors
t	0.00046***	0.00046***
	(0.00001)	(0.00002)
dt	-0.00075***	-0.00075***
	(0.00001)	(0.00005)
tue	0.1639***	0.1639***
	(0.01435)	(0.00956)
wed	0.20808***	0.20808***
	(0.01434)	(0.01141)
thu	0.2514***	0.2514***
	(0.01435)	(0.01081)
fri	0.21555***	0.21555***
	(0.01442)	(0.0108)
Constant	12.31408***	12.31408***
	(0.01386)	(0.02816)
Note:		*p<0.1; **p<0.05; ***p<0.01

The Newey-West standard errors of coefficients for both t and DT_t are greater than the uncorrected standard errors. This reflects the correction for autocorrelation that would otherwise inflate the significance of results. However, even after the correction t and DT_t are significant.

[10] These require stationarity, which was only suggested in the evidence of section 2.5, but will be assumed for the remaining sections.

2.6

The standard errors for days of the week decrease after the model is corrected for autocorrelation. Monday trading volume is *even more* significantly different from all other days (*i.e.*the results from Section 2.3 hold.) This result says that when we strip away any spurious autocorrelation, the days of the week effects were revealed to be stronger.

2.7

The eveidence presented in Sections 2.1 to 2.6 indicate a long-term effect due on the volume of shares traded on the S&P/ASX200 after the entrance of a competitor, Chi-X. Section 2.3 shows that the pre-Chi-X trend was 0.046%, reversing to -0.028% growth per trading day afterwards.