Room 1: Terrace

STAT 464/864 - Fall 2024

Discrete Time Series Analysis | Queen's University

Deadline for Original Problems: Friday, September 27th (midnight)

Galaxy Grading (Modified from *Specifications Grading,* Linda Nilson, 2014)

Instructions

This *Room* (assignment) has 6 *Stars* (problems). All are worth 5 points.

Choose any 4 to complete and submit. These are your **Original Stars.**
The remaining 2 are your **Alternate Stars.** You do not have to submit these.

Once you receive your grade for this *Room,* you can try for a higher score: Alternate Stars may be attempted *after* the Room's due date and, if favourable to your grade, may replace up to two of your Original Stars. All Alternate Stars are "due" December 2nd.

Graduate students must complete 4 stars, and also 1 **comet.** Yeah, there are 2 comets. Choose 1 Original Comet to submit – the other will be your Alternate Comet.

Fine Print

1. Alternate Stars may only replace Original Stars **from the same homework "Room."** This ensures that your improved grade reflects an improved understanding of the learning outcomes specific to that Room.
2. Only 2 Alternate Stars permitted per Room. (Ex: If you submit 3/4 Original Stars by the Room's due date, you can't use the missing Original Star as a 3rd Alternate Star.)
3. Comets (gradate problems) can be treated as Stars, but the converse is not true.

Learning Outcomes

1. Estimate trend and seasonal components of an observed time series, using R
2. Identify strong stationarity, weak stationarity, and non-stationary
3. Determine the ACVF and/or ACF of a stationary series
4. **[GRAD]** Derive limits (as $N \to \infty$) of sample ACFs for deterministic time-functions.

Formatting

Your submission **must** be a rendered Quarto document, otherwise it will not be graded. If you are having troubles rendering, please come to my office hours! I will help you!

R-based responses must use code chunks to execute code and display results. All written discussions must be typed in-document (for example: if you're asked to comment on the implications of a certain result, or to compare the performance of different models). See Workshop 1 for how to create a code chunk.

But you don't have to type out your *mathematical* proofs. You may write math formulae by hand, save the image as a **.png** in your **working directory** (the folder containing your **.qmd** file) and insert it into your document. The syntax to insert an image is:

`![Caption goes here](image.png){width=100%}`

"But Skye," you say, "isn't this just extra work? Since we upload it all to Crowdmark as individual pages, anyway?"

"Yes." I say. "But it helps ensure all your submissions are right-side-up, in order, etc. AND it gives a buffer of white space for the TA to leave comments. AND, it combines all the pieces of your homework submission and puts them in one place. AND this entire process is a good professional skill to make a habit of - it's built into RStudio because it's standard practice." Mic drop.

Beyond all this, there are a few basic expectations:

1. Don't print a bunch of unnecessary code. **Look** at the rendered pdf document. Is it 30 pages long? Are there hundreds of lines printing data values for no reason? In general, don't print out data. If you really want, you can give a preview of the data by using the command: `head(my_data)`.
2. Label your plots. Give each plot a relevant main title. The `x` axis will usually be time, in this course, but you should also indicate units and/or dates.
3. Organize your work using headers (hashtags). Example of syntax and output below.

    # Problem X

    #### Part y)
    Blablabla your answer here

Problem X

Part y)

Blablabla your answer here

Star 1: Carbon Dioxide Galaxy [R Problem]

Data	Global average marine surface CO_2 (parts per billion)
Times Sampled	(Monthly) January, 1980 – May, 2024
Source	NOAA (National Oceanic & Atmospheric Administration)

Load the dataset `co2.txt` into Rstudio. I recommend the code

```
# you can call the data variable [dat] whatever you want, of course
dat <- read.table("co2.txt", header = TRUE)
```

Part a) *Plotting*

Create a line plot of the time series, demonstrating the evolution of average marine surface CO_2 levels over time. **Comment** (as text in your Quarto doc) on any apparent trend and/or periodicity in the time series.

Plotting Etiquette

For full marks, hit the following criteria:

1. **Make it a line-plot.**
   This helps highlight the actual trajectory of the points.
   (You can see this in action if you try plotting a point-plot of the same data, and then compare the two plots. Which one better shows the trends we might be interested in? Not for marks, just food for thought.)
2. **Create an appropriate x-axis.**
   The time-units should be in *months,* however, the tick marks on the x-axis should correspond to *years.* (One tick every 12 observations.)
3. **Use scientifically meaningful labels:**
   Main title. x-axis. y-axis. Don't forget units!
4. **Don't panic:**
   Look, I know I'm really particular about the way I do things. But for YOU: all I care about are *learning outcomes.* As long as your plot is able to serve scientific purpose (which does require conveying details such as units and point trajectories), you've completed the mission. Beyond this, feel free to take risks, write your own functions (unless I say otherwise), use shortcuts, etc.

Part b) *Trend Estimation*

1. **Polynomial Regression:** Estimate the data's trend using a polynomial of degree $p = 2$.
2. **Moving Average Smoother:** Now estimate the trend using an moving average smoother with bandwidth $q = 10$.
3. **Plots:** Create a line-plot displaying your polynomial regression estimate on top of the original time series. Do the same for your MA smoother estimate in a separate plot. Don't forget your plotting etiquette (see part a).

   Be sure to use a different colours to distinguish the estimates from each other *and* the original series. The combo *black, red, and blue* is basically foolproof for R plots.
4. **Residuals:** Compute the residuals corresponding to each estimate. Display each resulting vector of residuals in its own line-plot.
5. **Comment:** which smooth do you think is better? **Why?**

Part c) *Seasonality*

Now that the trend has been eliminated, we can attend to the data's periodic structures. By "eliminated," I mean that you've calculated the residuals for your trend estimates: these residuals are just the original data with the (estimated) trend removed. In the following steps, you will be analysing the **residuals of the MA smoother's estimate** from part b.

1. **Harmonic Regression** In R, estimate a seasonal component s_t with period $d = 12$ (months), via harmonic regression.
2. **MA smoothing** Estimate the same s_t (still with $d = 12$) using ITSMR's `season()` function.
3. **Plots:** Create a line-plot displaying the harmonic regression fit on top of your MA smoother residuals from part b). Do the same for the `season()` fit, again plotted upon the residuals from part b). Don't forget your plotting etiquette.
4. **Residuals:** Compute the residuals corresponding to the `season()` fit. Make 2 plots: a line-plot of the residuals, and a plot of their ACF.
5. **Comment:** In part 3, why did the harmonic regression fit overestimate the peaks and valleys of s_t? Consider your plots from part 4. Is there any systematic seasonal structure remaining in these residuals? Why do you think so?

Star 2: Sad Wine Galaxy [R Problem]

> **Data** Sad Australian Red Wine Sales (unit = number of sales)
>
> **Times Sampled** (Quarterly: Mar/Jun/Sep/Dec) March, 1985 – June, 2014.
>
> **Source** A time-series-for-economics course I TA'd for, years ago.

Load the dataset `wine_sad.csv` into Rstudio. I recommend the code

```
dat <- read.csv("wine_sad.csv") # call the data [dat] whatever you want, tho
```

Part a) *Plotting*

Create a line plot of the time series, demonstrating the evolution of wine sales over time. **Comment** (as text in your Quarto doc) on any apparent trends/periodicity in the data.

Plotting Etiquette *See Carbon Dioxide Galaxy for more details*

For full marks, hit the following criteria:

1. Make it a line-plot
2. Create an appropriate x-axis
3. Use scientifically meaningful labels
4. Don't panic <3

Be sure to use a different colours to distinguish the estimates from each other *and* the original series. The combo *black, red, and blue* is basically foolproof for R plots.

Part b) *Trend Estimation*

1. **Moving Average Smoother:** Estimate the data's trend using an MA smoother with bandwidth $q = 2$.
2. **Exponential Smoother:** Now estimate the trend using an exponential smoother with smoothing parameter $\alpha = 0.5$.
3. **Plots:** Create a line-plot displaying your MA smoother estimate on top of the original time series. Do the same for your exponential smoother estimate in a separate plot. Don't forget your plotting etiquette.
4. **Residuals:** Compute the residuals corresponding to each estimate. Display each resulting vector of residuals in its own line-plot.
5. **Comment:** Which smooth do you think is better? **Why?**

Part c) *Seasonality*

Now that the trend has been eliminated, we can attend to the data's periodic structures. By "eliminated," I mean that you've calculated the residuals for your trend estimates: these residuals are just the original data with the (estimated) trend removed. In the following steps, you will be analysing the **residuals of the MA smoother's estimate** from part b.

1. **Harmonic Regression** In R, estimate a seasonal component s_t with period $d = 4$ (quarterly: 4 samples a year), via harmonic regression.
2. **Season Fitting** Estimate the same s_t (still with $d = 4$) using ITSMR's `season()` function.
3. **Plots:** Create a line-plot displaying the harmonic regression fit on top of your MA smoother residuals from part b). Do the same for the `season()` fit, again plotted upon the residuals from part b). Don't forget your plotting etiquette.
4. **Residuals:** Compute the residuals corresponding to the `season()` fit. Make 2 plots: a line-plot of the residuals, and a plot of their ACF.
5. **Comment:** In part 3, why did the harmonic regression fit overestimate the peaks and valleys of s_t?
   Consider your plots from part 4. Is there any systematic seasonal structure remaining in these residuals? Why do you think so?

Star 3: Moving Average Galaxy

Let $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ be a sequence of uncorrelated RVs. For all t,

1. Let $\mathrm{E}[\varepsilon_t] = 0$ and $\mathrm{Var}[\varepsilon_t] = 1$.
2. Let $X_t = \alpha_1\varepsilon_t + \alpha_2\varepsilon_{t-1}$ where α_1 and α_2 are non-zero, real constants.

Compute:

Derive the following in terms of time t, lag h, and the coefficients α_1 and α_2.

1. $\mathrm{Var}(X_t)$
2. $\mathrm{Cov}(X_t, X_{t+h})$
3. $\rho(X_t, X_{t+h})$

Comment:

Is the time series $\{X_t\}$ stationary?

Star 4: Functions-of-Noisy-Random-Variables Galaxy

Let $\{W_t\}$ be a zero-mean process of mutually independent RVs with variance σ^2. Consider the following time series, defined in terms of this W_t.

Show each is stationary, and compute its autocorrelation function.

Part a) $X_t = (-1)^t W_0 + (-1)^{t+1} W_1$

Part b) $X_t = W_0 \cos(\alpha t) + W_1 \sin(\alpha t)$

Where $\alpha \neq 0$ is a real-valued constant.

Hint: $\cos(x - y) = \cos x \cos y + \sin x \sin y$

Part c) $X_t = W_0 s_t + W_1 s_{t-1}$

Where s_t is seasonal, with period 2

Part d) $X_t = W_t W_{t-1} \cdots W_{t-j}$

Where j is a positive integer.

Star 5: Difference Operator Galaxy

Let $\{Z_t\}$ be a stationary time series, and let $\gamma_Z(h)$ denote its ACVF at lag h. Let s_t be seasonal with period $d = 10$. Let a and b be real valued constants.

Part a) *Let* $Y_t = a + bt + s_t + Z_t$.

Is $\{X_t \overset{def}{=} \nabla\nabla_{10}Y_t\}$ stationary? If so, compute its ACVF in terms of $\gamma_Z(h)$.

Part b) *Let* $Y_t = (a + bt)s_t + Z_t$

Is $\{X_t \overset{def}{=} \nabla_{10}^2 Y_t\}$ stationary? If so, compute its ACVF in terms of $\gamma_Z(h)$.

Star 6: Philosophical Galaxy

Part a)

In the heart of the Philosophical Galaxy, at the Time Series Cafe, your coffee date quin wants you to order her a time series that is weakly – but *not strongly* – stationary. What do you order for her? (Give an explicit formula). Why should she be happy with your choice?

Part b)

Let $\{W_t\}$ be an infinite, weakly stationary time series. Now consider the time series $\{Y_t \overset{def}{=} W_t^2\}$. Is $\{Y_t\}$ weakly stationary? If so, prove it. If not, find a counterexample.

Comet 1: Linear Time-Function Galaxy [GRAD Problem]

Let $\{y_t\}_{t=1}^{N}$ be an observed time series. Let $\hat{\rho}(h)$ be the lag-h sample autocorrelation function as defined in Definition 1.4.4 of Brockwell & Davis.

Suppose $x_t = a + bt$, where a and b are constants and $b \neq 0$. show that for each fixed $h \geq 1$,

$$\hat{\rho}(h) \overset{N\to\infty}{\longrightarrow} 1$$

Comet 2: Cosine Convergence Galaxy [GRAD Problem]

Let $x_t = c\cos(\omega t)$, where c and ω are constants: $c \neq 0$, and $0 \neq \omega \in (-\pi, \pi]$. Show that for $\forall\ h$,

$$\hat{\rho}(h) \overset{N\to\infty}{\longrightarrow} \cos(\omega h)$$

Hints :

$$\frac{\sin\Big((N+\frac{1}{2})\alpha\Big)}{\sin(\alpha/2)} = 1 + 2\cos(\alpha) + 2\cos(2\alpha) + 2\cos(N\alpha)$$

$$2\cos(\alpha)\cos(\beta) = \cos(\alpha-\beta) + \cos(\alpha+\beta)$$