Spring 2020, Math 545, Take-Home Final Exam

Due: 5PM PST on May 8th, 2020.

Instructions.

- You **can** use your book, notes, and previous homeworks on this exam.
- You **cannot use the internet or internet search resources** such as google, mathoverflow, etc. to complete the exam.
- You are required to show your work on each problem on the exam.
- For exercises involving coding, submit all of your code, and submit all output that is relevant to that exercise.
- The exam is **due at 5PM PST on May 8th, 2020**.
- You must submit the exam, via email, to `stevenmheilman@gmail.com`. This is the only email address at which the exam will be accepted.
- The exam must be submitted as a **single PDF file**. Submitting separate files will be penalized.
- **If you use a theorem or proposition from class or the notes or the book you must explicitly cite the theorem/proposition number from the book/notes** and explain why the theorem may be applied. (Since this particular exam is open note and open book, this extra requirement should be reasonable.)
- **Organize your work**, in a reasonably neat and coherent way. Work scattered all over the page without a clear ordering will receive very little credit.
- **Mysterious or unsupported answers will not receive full credit**. A correct answer, unsupported by calculations, explanation, or algebraic work will receive no credit; an incorrect answer supported by substantially correct calculations and explanations might still receive partial credit.

Exercise 1. Let $0 < \sigma < \infty$. Let $\{Z_n\}_{n\in\mathbf{Z}}$ be real-valued $WN(0,\sigma^2)$. Let $\{X_n\}_{n\in\mathbf{Z}}$ be a real-valued mean zero ARMA$(1,1)$ process defined by

$$X_n - \frac{1}{3}X_{n-1} = Z_n + \frac{1}{3}Z_{n-1}, \qquad \forall\, n \in \mathbf{Z}.$$

- Prove that there exist $c_0, c_1, \ldots \in \mathbf{R}$ such that

$$X_n = \sum_{j\geq 0} c_j Z_{n-j}, \qquad \forall\, n \in \mathbf{Z}.$$

  Also, explicitly find the constants $c_0, c_1, \ldots$.
- Write down an explicit formula for the autocovariance function $\gamma\colon \mathbf{Z} \to \mathbf{R}$ of $\{X_n\}_{n\in\mathbf{Z}}$.
- Find the value of the partial autocorrelation function $\alpha(1)$. Also, find the one-step prediction error $\mathbf{E}(X_2 - Y_2)^2$.
- Write down an explicit formula for the spectral density function $f\colon \mathbf{R}/\mathbf{Z} \to \mathbf{C}$ of $\{X_n\}_{n\in\mathbf{Z}}$.

Exercise 2 (Financial Data Analysis, Version 2). In this exercise, we will attempt to model financial data by ARMA processes. You can create a spreadsheet of financial data if you open up a spreadsheet in google spreadsheets, and input the following command into one of the cells in the spreadsheet (and then press enter):

```
=GOOGLEFINANCE("GOOG", "price", "1/1/2010", "12/31/2019", "DAILY")
```

The first argument in the command is the stock price ticker symbol (which corresponds to google stock). The closing price of the stock is then listed every day, starting from January 1st, 2010, and ending on December 31st, 2019. Once you have the spreadsheet data, you should be able to import it into Matlab or your favorite mathematical software.

A classic barometer of the US economy is the ten year US treasury note yield. You can access this data with the command

```
=GOOGLEFINANCE("TNX", "price", DATE(2005,1,1), DATE(2019,12,31), "DAILY")
```

That is, we are analyzing this data from 2005 to 2019. Let j be the (non-weekend) day occurring after January 1, 2005, so e.g. $j = 6$ corresponds to the sixth non-weekend day after January 1, 2005, and denote Y_j as the corresponding ten year US treasure note yield on day j. There should be $3,767$ entries in the data.

In a previous homework exercise, we examined the Yule-Walker estimators and autoregressive parameters for this data. In this exercise, we will try to find the trend and seasonal components of the data.

• In order to find the trend and seasonal components, first take the Fourier transform of the data, i.e. examine the function

$$\widehat{Y}(s) = \sum_{j=1}^{3767} Y_j e^{2\pi i s j}, \qquad \forall\, s \in \mathbf{R}/\mathbf{Z}$$

It should just have a large spike near zero, indicating a trend component, with perhaps no observed seasonal component. Plot the inverse Fourier transform of the large spike. That is, determine the frequencies s where the large spike is supported in the Fourier transform (say it is an interval of the form $[-a, a]$), and then plot

$$S_j := \int_{-a}^{a} \widehat{Y}(r) e^{-2\pi i r j} dr, \qquad \forall\, 1 \le j \le 3767.$$

Your choice of a should make S_j resemble the original data, but a should not be so large that it matches the small oscillations of the data too closely. (The exact choice of a might be a subjective choice.)

• The Fourier transform $\widehat{Y}(s)$ should look fairly "jagged." So, let's smooth it out to get a better picture of it (and verify whether or not a seasonal component exists). Recall that the

n^{th} **periodogram** of $Y_1, \ldots, Y_n$ is the function $I_n \colon \mathbf{R}/\mathbf{Z} \to [0, \infty)$ defined by

$$I_n(t) := \frac{1}{n} \left| \sum_{j=1}^{n} Y_j e^{2\pi itj} \right|^2, \qquad \forall\, t \in \mathbf{R}/\mathbf{Z}.$$

Define $w_n \colon \mathbf{Z} \to [0, \infty)$ so that $w_n(j) := c$ if $|j| \leq \sqrt{n}$, $w_n(j) := 0$ if $|j| > \sqrt{n}$, and $c > 0$ is chosen so that $\sum_{j \in \mathbf{Z}} w_n(j) = 1$. Recall that the **smoothed periodogram** is defined to be

$$I_{n,w}(t) := \sum_{j \in \mathbf{Z}} w_n(j) I_n(t - j/n), \qquad \forall\, t \in \mathbf{R}.$$

Plot the smoothed periodogram of the data. Does the result help you determine where the large spike is in the Fourier transform of the data? If so, amend the first part of the problem. Do you observe any seasonal component, i.e. a spike in the Fourier transform away from 0? If so, plot the seasonal component on its own.

• Finally, plot the data, with the trend (and seasonal component, if you found one) subtracted from the data. Call the resulting data $\{W_j\}_{j=1,2,\ldots,3767}$. We anticipate that $\{W_j\}_{j=1,2,\ldots,3767}$ should then be weakly stationary. Try to fit an autoregressive model $\mathrm{AR}(p)$ to $\{W_j\}_{j=1,2,\ldots,3767}$. That is, use the Yule-Walker estimators for the autoregressive parameters. Examine the values of the partial autocorrelation function, to try to determine a good value of p for the model. (Optional: use the AICC criterion to estimate the values of p, q in an $\mathrm{ARMA}(p, q)$ model for the data.)

• The (smoothed) periodogram of $\{W_j\}_{j=1,2,\ldots,3767}$ gives a good estimate for the spectral density of $\{W_j\}_{j=1,2,\ldots,3767}$. Using the MLARMA estimate (or any other estimate for the ARMA parameters of the process which are then plugged into an estimate of ϕ and θ, which then determine an estimate of the spectral density of the form $\sigma^2 |\theta/\phi|^2$), compare that estimate of the spectral density of $\{W_j\}_{j=1,2,\ldots,3767}$ to the smoothed periodogram of $\{W_j\}_{j=1,2,\ldots,3767}$. Do they appear similar to each other? To get a third comparison, plot the lag window spectral density estimator

$$F_{n,u}(t) := \sum_{j \in \mathbf{Z}} u(j/m) \Gamma_n(j) e^{2\pi itj}, \qquad \forall\, t \in \mathbf{R}/\mathbf{Z},$$

with $u(x) := 2(1 - |2x|) 1_{|x|<1/2}$ for any $x \in \mathbf{R}$, and $m_n := \lfloor \sqrt{n} \rfloor$ for all $n \geq 1$. Does this estimate of the spectral density resemble the other two? If not, adjust the parameters of the estimators, such as the smoothing function w_n for the smoothed periodogram, try changing the function m_n, or try other choices of u, as described in the book. Ideally, all of the estimators will be in agreement with each other, and they should agree with Theorem 8.9 in the notes, but in practice, this might not happen.

• Repeat all steps above for the logarithmically differenced google stock data from 2005 to 2019, i.e. using the command

```
=GOOGLEFINANCE("GOOG", "price", "1/1/2005", "12/31/2019", "DAILY")
```

That is, if X_n denotes the google stock price on day n, then consider

$$Y_n := \log(X_{n+1}/X_n) = \log X_{n+1} - \log X_n$$

and apply the above steps to Y_n. The behavior you examine here might be different than the behavior for the ten year treasury note yield.

Exercise 3 (Sunspot Data, Version 4). This exercise deals with sunspot data from the following files (the same data appears in different formats)

txt file   csv (excel) file

These files are taken from http://www.sidc.be/silso/datafiles#total

To work with this data, e.g. in Matlab you can use the command

```
x=importdata('SN_d_tot_V2.0.txt')
```

to import the .txt file.

The format of the data is as follows.

- Columns 1-3: Gregorian calendar date (Year, Month, then Day)
- Column 4: Date in fraction of year
- Column 5: Daily total number of sunspots observed on the sun. A value of -1 indicates that no number is available for that day (missing value).
- Column 6: Daily standard deviation of the input sunspot numbers from individual stations.
- Column 7: Number of observations used to compute the daily value.
- Column 8: Definitive/provisional indicator. A blank indicates that the value is definitive. A '*' symbol indicates that the value is still provisional and is subject to a possible revision (Usually the last 3 to 6 months)

In two previous exercises, we examined the number of sunspots U_t at time t, where t is measured in years. We also took the Fourier transform of U, and defined

$$\widehat{U}(r) := \sum_{t \in \mathbf{Z}/365} U_t e^{2\pi i t r}, \qquad \forall\, r \in \mathbf{R}/365\mathbf{Z}.$$

We found that the seasonal component of U closely matched the following function of $t \in \mathbf{Z}/365$

$$S_t := \frac{1}{365} \int_{.08}^{.105} \widehat{U}(r) e^{-2\pi i r t} dr + \frac{1}{365} \int_{-.105}^{-.08} \widehat{U}(r) e^{-2\pi i r t} dr.$$

We also found that the trend component of U closely matched the following function of $t \in \mathbf{Z}/365$

$$M_t := \frac{1}{365} \int_{-.016}^{.016} \widehat{U}(r) e^{-2\pi i r t} dr.$$

In this exercise, we will smooth out the Fourier transform to get a better picture of it.

- Plot the periodogram and smoothed periodogram of the sunspot data (using the same smoothing function w_n as in the previous Exercise). Compare the two plots. Does the

smoothed periodogram have a spike around frequency $1/11$? What is the exact frequency of the highest value of the spike in the Fourier transform?

• The (smoothed) periodogram of $U_t - S_t - M_t$ gives a good estimate for the spectral density of $U_t - S_t - M_t$. Using the MLARMA estimate (or any other estimate for the ARMA parameters of the process which are then plugged into an estimate of ϕ and θ, which then determine an estimate of the spectral density of the form $\sigma^2 |\theta/\phi|^2$), compare that estimate of the spectral density to the smoothed periodogram. Do they appear similar to each other? To get a third comparison, plot the lag window spectral density estimator

$$F_{n,u}(t) := \sum_{j \in \mathbf{Z}} u(j/m)\Gamma_n(j)e^{2\pi itj}, \qquad \forall\, t \in \mathbf{R}/\mathbf{Z},$$

with $u(x) := 2(1 - |2x|)1_{|x|<1/2}$ for any $x \in \mathbf{R}$, and $m_n := \lfloor \sqrt{n} \rfloor$ for all $n \geq 1$. Does this estimate of the spectral density resemble the other two? If not, adjust the parameters of the estimators, such as the smoothing function w_n for the smoothed periodogram, try changing the function m_n, or try other choices of u, as described in the book. Ideally, all of the estimators will be in agreement with each other, and they should agree with Theorem 8.9 in the notes, but in practice, this might not happen.