2017 PSS SUMMER SCHOOL
Assignment 2a

Due: 12:30pm on June 23

1 Submission instructions

To submit your assignment, email all of your files in a single zip folder to pss.ss.hw.submit@sp.frb.gov. Please using the naming convention <GIVENNAME>_<SURNAME>_a<NUMBER>_m1_2017.<FILEXTENSION> for any **file that you edit or create**, for files that you simply use without modifying, such as the assignment directions or a data file, leave the name as is, usually something like a1_m1_2017.pdf. The below directory comparison shows you an example of the zip directory you will receive and the one you will email submit after completion.

Note: Assignment 2 has two parts, a and b, this assignment is part a. Even though there are two parts you will continue use the same submission folder for both, simply adding to it for the second assignment. Name this directory <GIVENNAME>_<SURNAME>_a2_m1_2017.zip.

(zip to students)

```
a1_m1_2017/
├─ a1_m1_2017.pdf
├─ code/
├─ data/
│  ├─ input/
│  │  └─ a1_m1_2017.Rdata
│  └─ output/
└─ plots/
```

(zip from students)

```
<GIVENNAME>_<SURNAME>_a1_m1_2017/
├─ a1_m1_2017.pdf
├─ <GIVENNAME>_<SURNAME>_a1_m1_2017.pdf
├─ code/
│  ├─ <GIVENNAME>_<SURNAME>_a1_m1_2017.R
│  └─ <GIVENNAME>_<SURNAME>_a1_m1_2017.tex
├─ data/
│  ├─ input/
│  │  └─ a1_m1_2017.Rdata
│  └─ output/
│     └─ <GIVENNAME>_<SURNAME>_a1_m1_2017.Rdata
└─ plots/
   ├─ <GIVENNAME>_<SURNAME>_a1_m1_2017.png
   └─ <GIVENNAME>_<SURNAME>_a1_m1_2017.jpg
```

Where you make the proper substitutions, e.g. JUSTIN_SKILLMAN_a3_m1_2017.zip. Remember that you don't need to change the names of the files you do not modify, such as a1_m1_2017.Rdata, only the files that you create or modify. Your code for each assignment example would be saved under code/ in a .R file, named as shown above.

Input/Output from R

Below are some basic commands you should put in your script

```
load(file=[filename])
[fullpathfilename] <- paste0([path_to_file],[filename])
save([variables]*, file=[fullpathfilename])
```

For future reference:

```
df <- read.csv([filename])
write.csv(df, file=[filename])

library(foreign)
df <- read.dta([filename])
write.dta(df, file=[filename])
```

List Submission

Please submit all answers in a list data structure. Below is a basic example, instead of *result* use the name given in each problem:

```
> answers <- list()
> result <- 10 + 15
> answers$q1a <- result

> answers$q1a
[1] 25
```

We start out by generating an empty list to eventually store all answers in. Now let us say that we calculated the answer to question q1a and stored it in the variable result. We then add the result value to the answers list as shown above.

2 Estimating Macroeconomic Variables

This question will use most of what you have learned up to this point in the module. Remember that you shouldn't view the problem in its entirety but rather break it down into smaller parts that are more manageable to approach. We assist you in this, leading you through the problem incrementally. The dataset `agg_econ` is from FRED and contains the following aggregate economic variables, we observe 68 years 1949:2016.

Variable name	**Description**	**Desired movement**
`gdp_chng`	% Change in gross domestic product (annual)	$\uparrow$
`exprt_chng`	% Change in net trade exports (annual)	$\uparrow$
`unemp_chng`	% Change in unemployment (annual)	$\downarrow$
`prices_chng`	% Change in consumer prices (annual)	$\uparrow$

We are proposing a model that aims to predict whether GDP will increase or decrease in the next year given knowledge of how our other variables have moved in past years. The model we are working with looks like,

$$\hat{G} = \gamma_1 T_{k_1} + \gamma_2 U_{k_2} + \gamma_3 P_{k_3}$$

where T, U, and P are scalar binary variables referring to the behavior of `exprt_chng`, `unemp_chng`, and `prices_chng` respectively. We denote our independent variables as $\vec{X} = \{T, U, P\}$. $\hat{G}$ is our dependent variable, it is *not* binary and can take any real value $[0, 1]$. Our observed G can only be zero or one depending on if GDP decreased or increased respectively. The index vector $\vec{k} = \{k_1, k_2, k_3\}$ represents the number of past years over which we will determine if a majority of the annual changes were desired changes in these variables. Majority defined as greater than half ($\frac{k_i}{2}$) of the observations. The vector $\vec{\gamma} = \{\gamma_1, \gamma_2, \gamma_3\}$ represents the weights we put on each variable, its elements can be any real value. As an example if the data showed `unemp_chng=` `{.05,-.01,.04,-.12}` from 2013 through 2016 (we are at year-end of 2016), U_{k_2} (where $k_2 = 3$) would be **1** because there were two decreases (which are desired for unemployment rates) out of the last three annual changes. U_{k_2} (where $k_2 = 4$) however would be **0** because only two of the last four changes has been desired. $\hat{G}$ represents our prediction of the change in GDP, values closer to 1 indicate a stronger prediction of an increase. Please note that $\hat{G}$ must be between 0 and 1, which means that your model is subject to the constraint that the $\vec{X} \cdot \vec{\gamma} \in [0, 1]$ ($\cdot$ is the dot product).

The goal of the model and optimization is to predict if G is 0 or 1 in the next year, representing a decrease or increase in GDP respectively. To do this we need to determine *both* what the optimal γ's and set of k's are. Optimality is defined as giving the best predictions on a training set, where we minimize the difference between our predicted $\hat{G}$ and observed G. We will lead you through the beginning stages of reaching this goal but the final step requires completing the bonus section which is not mandatory but is encouraged!

Problem 2.1. Create a function to compute T, U, and P for a given $\vec{k}$ and single year. For your answer assume $\vec{k} = \{8, 4, 6\}$ and use 2012 as your current year. This means that your results, `chngs_846` which will be saved in `answers`, will be a vector of 1's and 0's of length three. Save the function in your code but you do not need to add it to the `answers` data.

Problem 2.2. Use `chngs_846` as your $\vec{X}$ to compute a solution (there are multiple) for $\vec{\gamma}$ where each element is non-zero and < 1. Your observed G will correspond to GDP's change in 2013, as if you were trying to predict it only knowing data from 2012. To do this, minimize in the mean squared sense the difference between $\hat{G}$ and the observed data G. Create a function to find this $\vec{\gamma}$ as it will be used later. Save this vector as `weights_given`, similarly this will be saved in `answers`. This minimization would look like,

$$\min_{\vec{\gamma}} \frac{\sum_{i=1}^{n} (\hat{G} - G)^2}{n}.$$

Please note that $n = 1$ here as we are only working with one year's data. In later problems n will be larger.

HINT: Modularize your function so that the function returning the mean squared difference is called within the optimization funtion. Use an optimization routine such as the `constrOptim` in the `stats` package. When using these you need to specify an initial "guess", θ, choose anywhere in the solution region. The constraint setup is written below, you will still need to translate it to R. A is the constraint matrix, x is what we are searching for, and b is our constraint vector.

$$Ax \geq b$$

$$\begin{pmatrix} T & U & P \\ -T & -U & -P \end{pmatrix} \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \end{pmatrix} \geq \begin{pmatrix} 0 \\ -1 \end{pmatrix}$$

Problem 2.3. Using $\vec{k} = \{8, 4, 6\}$ as above, compute your $\vec{X} = \{T, U, P\}$ for all periods in the data, with the exception of 2016 as we don't have a 2017 G to test against. Remove any years for which not every element in $\vec{X}$ can be computed. Save this as a dataframe, of dimension 60×3, as `X_mtx` in `answers`. For notational reasons we will refer to this matrix (set of $\vec{X}$'s) as X_M.

Problem 2.4. Compute X_M dataframes for each of the five $\vec{k}$'s shown below, over the appropriate years for each $\vec{k}$. Determine which $\vec{k}$ leads to the best prediction of G given a constant $\vec{\gamma} = \{\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\}$, again by "best" in the mean squared sense. This will be computed over all the rows of X_M not just one, so $\hat{G}$ and G will be vectors spanning the appropriate observation time. Save only the $\vec{k}$ that results in the best prediction as `k_bestof5` in `answers`.

$$\vec{k}^A = \{1, 1, 1\}$$
$$\vec{k}^B = \{2, 2, 3\}$$
$$\vec{k}^C = \{12, 3, 4\}$$
$$\vec{k}^D = \{5, 7, 9\}$$
$$\vec{k}^E = \{6, 3, 4\}$$