

# Graduate Econometrics Recitations: 2021-2022

Alex Houtz\*

March 24, 2023

# Contents

<b>Preface</b>	<b>4</b>
<b>1 Introduction to Matlab</b>	<b>7</b>
1.1 Matlab Basics . . . . .	7
1.2 Matlab Set-up . . . . .	8
1.3 Some Useful Commands . . . . .	8
1.4 Publishing . . . . .	9
1.5 General Advice . . . . .	10
<b>2 Univariate Probability</b>	<b>11</b>
2.1 Monte Hall . . . . .	11
2.2 Practice Problem 1: Hansen 2.7 . . . . .	13
2.3 Practice Problem 2: Hansen 2.11 . . . . .	14
2.4 Practice Problem 3: Hansen 3.5 . . . . .	14
2.5 Practice Problem 4: Hansen 3.9 . . . . .	17
<b>3 Multivariate Probability</b>	<b>19</b>
3.1 Expected Value of a Log-Normal Distribution . . . . .	19
3.2 Practice Problem 1: Hansen 4.1 . . . . .	20
3.3 Practice Problem 2: Hansen 4.7 . . . . .	23
3.4 Practice Problem 3: Multivariate Change of Variables . . . . .	23
<b>4 Dependent Random Variables</b>	<b>27</b>
4.1 Previous Problem: Hansen 4.14 . . . . .	27
4.2 Practice Problem 1: Hansen 4.19 . . . . .	28
4.3 Practice Problem 2: AR(2) Process . . . . .	29
4.4 Practice Problem 3: Deriving the MA( $\infty$ ) Form for an AR(1) . . . . .	33
<b>5 Dependent Vectors of Random Variables</b>	<b>35</b>
5.1 Previous Problem 1: Hansen 5.12 . . . . .	35
5.2 Previous Problem 2: Deriving the F Distribution . . . . .	36
5.3 Practice Problem 1: Companion Form . . . . .	38
5.4 Matlab Help . . . . .	42
<b>6 Bias and Consistency</b>	<b>45</b>

6.1	The Analogy Principle . . . . .	45
6.2	Bias . . . . .	45
6.3	Consistency . . . . .	48
6.4	Midterm Review . . . . .	49
<b>7</b>	<b>Asymptotics</b>	<b>53</b>
7.1	Previous Problem: Gaussian Autoregressive Process . . . . .	53
7.2	Practice Problem 1: Hansen 8.1 . . . . .	56
7.3	Practice Problem 2: Hansen 8.3 . . . . .	58
7.4	Practice Problem 3: Hansen 8.8 . . . . .	58
7.5	Matlab Checkpoints . . . . .	60
<b>8</b>	<b>Maximum Likelihood Estimation</b>	<b>63</b>
8.1	Maximum Likelihood Theory . . . . .	63
8.2	Practice Problem 1: Hansen 10.4 . . . . .	65
8.3	Practice Problem 2: Hansen 10.7 . . . . .	66
8.4	Practice Problem 3: Hansen 10.8 . . . . .	67
8.5	Matlab Help . . . . .	69
<b>9</b>	<b>Method of Moments</b>	<b>73</b>
9.1	Previous Problem 1: Matlab Theory Part . . . . .	73
9.2	Method of Moments Theory . . . . .	77
9.3	Practice Problem 1: Hansen 11.3 . . . . .	78
9.4	Practice Problem 2: Hansen 11.4 . . . . .	79
9.5	Matlab Help . . . . .	79
<b>10</b>	<b>Hypothesis Testing and Confidence Intervals</b>	<b>83</b>
10.1	Previous Problem 1: OLS Estimator . . . . .	83
10.2	Hypothesis Testing . . . . .	85
10.3	Other Tests . . . . .	86
10.4	Practice Problem 1: Hansen 13.3 Extended . . . . .	87
10.5	Some Matlab Help . . . . .	90
<b>11</b>	<b>Basics of Regression</b>	<b>91</b>
11.1	Previous Problem 1: Hansen 13.1 Extended . . . . .	91
11.2	Regression Theory . . . . .	94
11.3	Matlab Help . . . . .	100
11.4	Previous Problem 2: Hansen II 3.13 . . . . .	101
11.5	Practice Problem 1: Hansen II 4.16 Adapted . . . . .	104
11.6	Practice Problem 2: Hansen II 7.1 . . . . .	106
11.7	Practice Problem 3: Hansen II 7.15 . . . . .	107
11.8	Practice Problem 4: Hansen II 7.23 Adapted . . . . .	108

<b>12 Summary of Econometrics I</b>	<b>111</b>
12.1 Previous Problem: Hansen II 7.7 . . . . .	111
12.2 Maximum Likelihood Estimation: Hansen 10.8 . . . . .	113
12.3 Method of Moments Estimation: Hansen 11.3 . . . . .	116
12.4 Regression Tests: Hansen 13.3 Extended . . . . .	117
12.5 Final . . . . .	119

# Preface

This book is a compilation of recitations given throughout the 2021-22 academic year at the University of Notre Dame for the first-year PhD econometrics sequence in the Department of Economics. They were originally composed by Alex Houtz, who was the graduate teaching assistant for Drew Creal in Fall 2021 and Marinho Bertanha in Spring 2022. The materials within may be distributed at any time to any audience for individual or instructional use, given proper citation to the author. These materials may not be used for commercial purposes.

Many problems were taken or adapted from Bruce Hansen's two econometrics books, which can be found [here](#). The first book is referred to as simply "Hansen" while the second book is referred to as "Hansen II." Other textbooks that are drawn from include Wooldridge (found [here](#)) and Hayashi (found [here](#)). The remaining problems are either cited in-text or are taken from the homework given in the lecture class by either Drew Creal or Marinho Bertanha.

The following list documents the chain of teaching assistants at the University of Notre Dame that used and/or edited this book, from earliest to latest:

1. Alex Houtz (2021-2022)

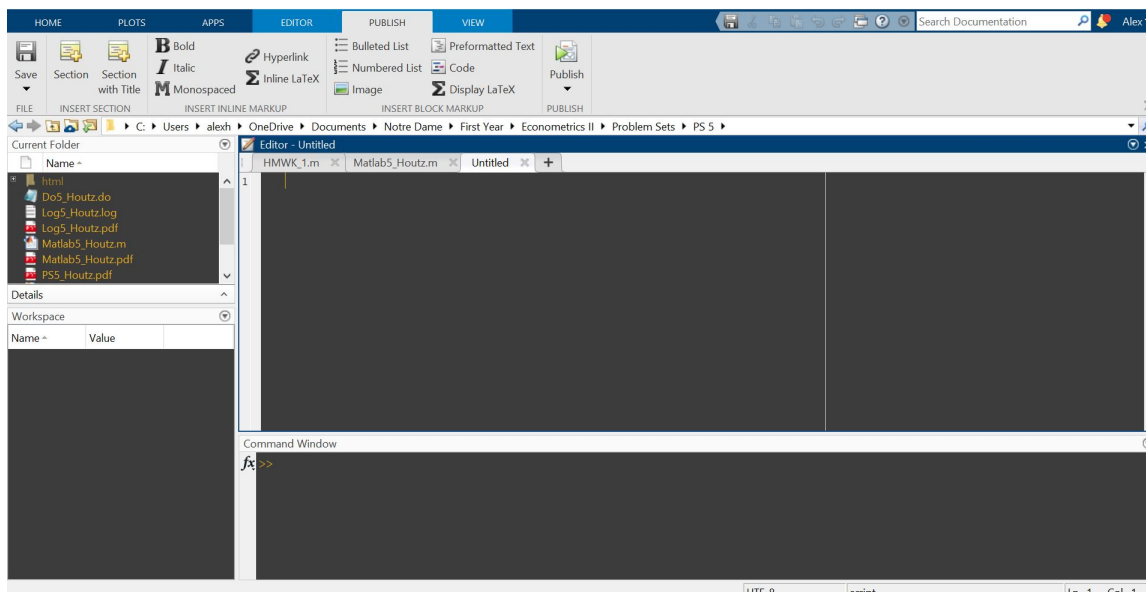


# Chapter 1

## Introduction to Matlab

### 1.1 Matlab Basics

We need to begin with the structure of Matlab.



- The "Editor" window (upper right) is where we create the code file. Your saved work will come from here.
- The "Command" window (bottom right) contains output from your code. You can also directly input code here that you don't want saved.
- The "Current Folder" window (upper left) contains the files specified in the path I have declared (see below in Matlab Set-up).

- The "Workspace" window (bottom left) contains saved variables, matrices, arrays, etc.

## 1.2 Matlab Set-up

There will be many Matlab assignments throughout the year. Establishing a simple, clean set-up to your code will make all of our lives easier. Feel free to experiment and create your own, but here is an example:

```
%% Introduction
% Alex Bouts
% Econometrics II
% Problem Set 5

%% Settings
clear, clc
cd('C:\Users\alexh\OneDrive\Documents\Notre Dame\Econometrics II\Problem Sets\PS 5');
```

There are a couple "tricks" to note:

- The double percent signs, "%%", create new sections. So in the example, I have the section "Introduction" and the section "Settings".
- The single percent sign, "%", tells Matlab that the following line is a comment and not code.
- The "clear, clc" commands clear the workspace and therefore ensure I'm not overwriting other variables in other files.
- The "cd(...)" command specifies the path on which Matlab will look for data and function files.
- The semi-colon, ";" suppresses output in the command window.

## 1.3 Some Useful Commands

```
%% Examples
% Define a variable
x = 10000;

% Import Data (can also use readtable)
cd('C:\Users\alexh\OneDrive\Documents\Notre Dame\First Year\Econometrics I\Homework\Homework 1');
data = readmatrix("cps09mar.xlsx");

% For loop
mu = [0 0];
Sigma = [1 .9; .9 1];

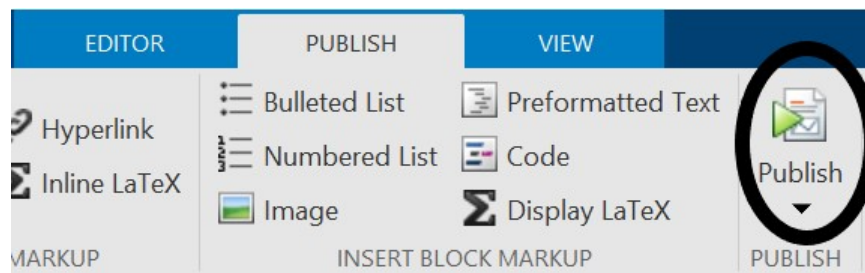
for g = 1:10000
    R = mvnrnd(mu, Sigma, 100);
    u(:,g) = R(:,1);
    y(:,g) = R(:,2);
end
```



- Variables must be defined for values, function outputs, and data if you want to reference them later in your code.
- To import data, declare your path then read the data into Matlab using "readmatrix" or "readtable". See documentation for more details on these commands.
- For loops will be useful in most Matlab assignments. In the example above, I'm creating two error term vectors with a multivariate normal distribution. Essentially, Matlab is pulling two random numbers from the distribution 10,000 times. NOTE: Try to avoid triple loops if possible. These can take a long time to run.

## 1.4 Publishing

Your code needs to be able to run from the beginning to the end with no problems. One easy way to check this, and also to turn in your code and output, is to publish your code.



- Usually we work in the Editor tab. To publish, click the Publish tab and then click on Publish.
- When the preview pops up, print to PDF. The end result should look similar to the picture below
- Make sure the necessary output is visible in the published PDF. Remember that by using a semi-colon you suppress the output.

**Contents**

---

- [Introduction](#)
- [Settings](#)
- [Part a](#)
- [Part b](#)
- [Part c](#)
- [Part d](#)
- [Part e](#)
- [Part f](#)
- [Part g](#)

**Introduction**

---

```
% Alex Houtz  
% Econometrics II  
% Problem Set 6
```

---

**Settings**

---

```
clear, clc  
cd('C:\Users\alexh\OneDrive\Documents\Notre Dame\Econometrics II\Problem Sets\PS 6');  
  
format short g;  
data=load('draft_lott.txt');
```

---

## 1.5 General Advice

- Coding is hard. Do not worry if you are struggling. Find a friend to code with and work through these together (do make sure you understand what the code is doing).
- Try to reduce run times for your code. Most assignments should be able to run in less than a minute.
- The internet is your best friend for Matlab (and code in general). Check out places like Matlab Answers first.
- Make sure to save your work.

## Chapter 2

# Univariate Probability

### 2.1 Monte Hall

This problem is important. Last year we had a question on the midterm using the same principles found in the Monte Hall problem. In light of this information, I will lay out the reasoning for this problem and solve it below. Make sure you know how to solve this question and any like it.

You are on the game show "Let's Make a Deal with Monte Hall." Without loss of generality, you pick door A but it is not immediately opened. To increase the drama, Monte opens one of the two remaining doors, door B (again, without loss of generality) revealing that door B does not hide the grand prize. Monte makes you an offer: Either switch your door choice to C or keep door A. Should you switch doors?

**Solution:** We begin by first defining variables.

$P \equiv$  Door hiding the prize

$S \equiv$  Door selected by the player

$H \equiv$  Door opened by the host

We begin with the assumption that, WLOG, you select door A. As such, the probabilities for the location of the prize are:

$$\mathbb{P}(P = 1|S = 1) = \frac{1}{3}$$

$$\mathbb{P}(P = 2|S = 1) = \frac{1}{3}$$

$$\mathbb{P}(P = 3|S = 1) = \frac{1}{3}$$

Using Bayes' Rule, we know that

$$\mathbb{P}(P = 1|H = 2, S = 1) = \frac{\mathbb{P}(H = 2|P = 1, S = 1)\mathbb{P}(P = 1|S = 1)}{\mathbb{P}(H = 2|S = 1)}$$

and

$$\mathbb{P}(P = 3|H = 2, S = 1) = \frac{\mathbb{P}(H = 2|P = 3, S = 1)\mathbb{P}(P = 3|S = 1)}{\mathbb{P}(H = 2|S = 1)}$$

Note that we are missing the components of the total probability of  $\mathbb{P}(H = 2|S = 1)$ . But we know these from your perspective:

$$\mathbb{P}(H = 2|P = 1, S = 1) = \frac{1}{2}$$

$$\mathbb{P}(H = 2|P = 2, S = 1) = 0$$

$$\mathbb{P}(H = 2|P = 3, S = 1) = 1$$

We can now compute the probabilities we want:

$$\begin{aligned} \mathbb{P}(P = 1|H = 2, S = 1) &= \frac{\mathbb{P}(H = 2|P = 1, S = 1)\mathbb{P}(P = 1|S = 1)}{\mathbb{P}(H = 2|S = 1)} \\ &= \frac{\frac{1}{2} \times \frac{1}{3}}{\frac{1}{2} \times \frac{1}{3} + 0 \times \frac{1}{3} + 1 \times \frac{1}{3}} \\ &= \frac{\frac{1}{2}}{\frac{1}{2} + 1} \\ &= \frac{1}{3} \end{aligned}$$

Where  $\mathbb{P}(H = 2|S = 1) = \sum_{i=1}^3 \mathbb{P}(H = 2|P = i, S = 1)\mathbb{P}(P = i|S = 1)$  via the law of total probability.

Applying the same steps:

$$\begin{aligned}
\mathbb{P}(P = 3|H = 2, S = 1) &= \frac{\mathbb{P}(H = 2|P = 3, S = 1)\mathbb{P}(P = 3|S = 1)}{\mathbb{P}(H = 2|S = 1)} \\
&= \frac{1 \times \frac{1}{3}}{\frac{1}{2} \times \frac{1}{3} + 0 \times \frac{1}{3} + 1 \times \frac{1}{3}} \\
&= \frac{\frac{1}{3}}{\frac{1}{6} + \frac{1}{3}} \\
&= \frac{2}{3}
\end{aligned}$$

Therefore, the probability the prize is behind your first choice, door A, is  $\frac{1}{3}$  while the probability the prize is behind your alternative choice, door C, is  $\frac{2}{3}$ . Given that you want to win, you should switch your door choice to C.

## 2.2 Practice Problem 1: Hansen 2.7

We are given the chi-square distribution. We are tasked with deriving the pdf for the inverse chi-square distribution. The pdf and change equation are given below:

$$\begin{aligned}
f_x(x) &= \frac{1}{2^{r/2}\Gamma(\frac{r}{2})} x^{r/2-1} e^{-x/2} \\
Y &= 1/X
\end{aligned}$$

Therefore, we need the change of variables formula:

$$f_y(y) = \left| \frac{\partial g^{-1}(y)}{\partial y} \right| f_x(g^{-1}(y))$$

First we find  $g^{-1}(y)$ . We know  $g(x) = Y = 1/X$ . So  $g^{-1}(y) = 1/Y$ .

Then we plug this into the chi-square pdf:

$$f_x(g^{-1}(y)) = \frac{1}{2^{r/2}\Gamma(\frac{r}{2})} \left(\frac{1}{y}\right)^{r/2-1} e^{-1/(2y)}$$

Finding the Jacobian next:

$$\begin{aligned} \left| \frac{\partial g^{-1}(y)}{\partial y} \right| &= \left| \frac{-1}{y^2} \right| \\ &= \frac{1}{y^2} \end{aligned}$$

Lastly, putting it all together:

$$\begin{aligned} f_y(y) &= y^{-2} \frac{1}{2^{r/2} \Gamma(\frac{r}{2})} y^{-r/2+1} e^{-1/(2y)} \\ &= \frac{1}{2^{r/2} \Gamma(\frac{r}{2})} y^{-r/2-1} e^{-1/(2y)} \end{aligned}$$

## 2.3 Practice Problem 2: Hansen 2.11

Suppose  $X$  has density  $f(x) = e^{-x}$ . Set  $Y = -\ln(X)$  and find  $f_y(y)$ .

To solve, we follow the same procedure as before.

First we find  $g^{-1}(y)$ . We know  $g(x) = Y = -\ln(X)$ . So  $g^{-1}(y) = e^{-y}$ .

Plugging this into the pdf:

$$f_x(g^{-1}(y)) = e^{-e^{-y}}$$

Then finding the Jacobian:

$$\begin{aligned} \left| \frac{\partial g^{-1}(y)}{\partial y} \right| &= |-e^{-y}| \\ &= e^{-y} \end{aligned}$$

Now altogether:

$$f_y(y) = e^{-y} e^{-e^{-y}}$$

## 2.4 Practice Problem 3: Hansen 3.5

For the exponential distribution, show:

- (a)  $\int_0^\infty f(x|\lambda) dx = 1$
- (b)  $\mathbb{E}[X] = \lambda$

$$(c) \operatorname{var}(X) = \lambda^2$$

### 2.4.1 Part a

Hansen defines the exponential distribution as follows:

$$f_x(x) = \frac{1}{\lambda} e^{-\frac{x}{\lambda}}$$

Plug this into the integral and solve:

$$\begin{aligned} \int_0^{\infty} \frac{1}{\lambda} e^{-\frac{x}{\lambda}} dx &= \frac{1}{\lambda} \int_0^{\infty} e^{-\frac{x}{\lambda}} dx \\ &= \frac{1}{\lambda} \left[ -\lambda e^{-\frac{x}{\lambda}} \right]_0^{\infty} \\ &= \frac{1}{\lambda} [-\lambda(0 - 1)] \\ &= \frac{\lambda}{\lambda} \\ &= 1 \quad \blacksquare \end{aligned}$$

### 2.4.2 Part b

Using the definition of the expected value  $\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_x(x) dx$ :

$$\begin{aligned} \mathbb{E}[X] &= \int_0^{\infty} \frac{x}{\lambda} e^{-\frac{x}{\lambda}} dx \\ &= \frac{1}{\lambda} \int_0^{\infty} x e^{-\frac{x}{\lambda}} dx \end{aligned}$$

We need to do integration by parts. So by defining:

$$\begin{aligned} u &= x & dv &= e^{-\frac{x}{\lambda}} dx \\ du &= dx & v &= -\lambda e^{-\frac{x}{\lambda}} \end{aligned}$$

we can substitute into the integration by parts formula:

$$\int u dv = uv - \int v du$$

Now we have

$$\frac{1}{\lambda} \left\{ \left[ -x\lambda e^{-\frac{x}{\lambda}} \right]_0^{\infty} + \lambda \int_0^{\infty} e^{-\frac{x}{\lambda}} dx \right\}$$

We use L'Hopital's Rule to evaluate the first term at the upper limit:

$$\begin{aligned} \lim_{x \rightarrow \infty} -x\lambda e^{-\frac{x}{\lambda}} &= \lim_{x \rightarrow \infty} \frac{-x\lambda}{e^{\frac{x}{\lambda}}} \\ &= \lim_{x \rightarrow \infty} \frac{-x\lambda}{\frac{1}{e^{-\frac{x}{\lambda}}}} \end{aligned}$$

This puts the expression in an  $\infty/\infty$  form. Taking the derivative of the top and bottom:

$$\lim_{x \rightarrow \infty} \frac{-\lambda}{\frac{1}{\lambda}e^{\frac{x}{\lambda}}} = 0$$

Back to the problem:

$$\begin{aligned} \frac{1}{\lambda} \left\{ [-x\lambda e^{-\frac{x}{\lambda}}]_0^\infty + \lambda \int_0^\infty e^{-\frac{x}{\lambda}} dx \right\} &= \frac{1}{\lambda} \left\{ [0 - 0] + \lambda [-\lambda e^{-\frac{x}{\lambda}}]_0^\infty \right\} \\ &= \frac{\lambda^2}{\lambda} \\ &= \lambda \quad \blacksquare \end{aligned}$$

### 2.4.3 Part c

Recall the variance decomposition formula:  $\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$ . We know  $\mathbb{E}[X]$ , so all we need is  $\mathbb{E}[X^2]$ . The steps are very similar as above:

$$\mathbb{E}[X^2] = \frac{1}{\lambda} \int_0^\infty x^2 e^{-\frac{x}{\lambda}} dx$$

We need to do integration by parts again. Define

$$\begin{aligned} u &= x^2 & dv &= e^{-\frac{x}{\lambda}} dx \\ du &= 2x dx & v &= -\lambda e^{-\frac{x}{\lambda}} \end{aligned}$$

Continuing on by plugging these into the integration by parts formula:

$$= \frac{1}{\lambda} \left\{ [-\lambda x^2 e^{-\frac{x}{\lambda}}]_0^\infty + 2\lambda \int_0^\infty x e^{-\frac{x}{\lambda}} dx \right\}$$

We can use L'Hopital's Rule to show that the first term goes to 0 again. Even better, the second term is the same term we solved when finding  $\mathbb{E}[X]$ ! Substituting everything in:



$$\begin{aligned}
 &= \frac{1}{\lambda}[0 + 2\lambda^3] \\
 &= 2\lambda^2
 \end{aligned}$$

Plugging  $\mathbb{E}[X^2]$  and  $\mathbb{E}[X]$  into the variance decomposition formula, we get:

$$\begin{aligned}
 \text{Var}(X) &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 \\
 &= 2\lambda^2 - \lambda^2 \\
 &= \lambda^2 \quad \blacksquare
 \end{aligned}$$

## 2.5 Practice Problem 4: Hansen 3.9

For the gamma distribution, show:

$$(a) \int_0^\infty f(x|\alpha, \beta) dx = 1$$

$$(b) \mathbb{E}[X] = \frac{\alpha}{\beta}$$

$$(c) \text{var}(X) = \frac{\alpha}{\beta^2}$$

### 2.5.1 Part a

We begin by noting that the pdf of a gamma distribution is  $f_x(x|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$ . Subbing this into the integral:

$$\int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} dx$$

Using property 3 in Hansen Appendix A.28 ( $\int_0^\infty x^{\alpha-1} e^{-\lambda x} dx = \frac{\Gamma(\alpha)}{\lambda^\alpha}$ ), we get:

$$\begin{aligned}
 &= \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot \frac{\Gamma(\alpha)}{\beta^\alpha} \\
 &= 1 \quad \blacksquare
 \end{aligned}$$

### 2.5.2 Part b

Use the definition for the expected value of a random variable:

$$\begin{aligned}
\mathbb{E}[X] &= \int_0^{\infty} \frac{\beta^\alpha}{\Gamma(\alpha)} x^\alpha e^{-\beta x} dx \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot \frac{\Gamma(\alpha + 1)}{\beta^{\alpha+1}} && (\text{Prop. 3}) \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot \frac{\alpha \Gamma(\alpha)}{\beta^{\alpha+1}} && (\text{Prop. 2}) \\
&= \frac{\alpha}{\beta} \quad \blacksquare
\end{aligned}$$

### 2.5.3 Part c

Using the variance decomposition equation, we again only need  $\mathbb{E}[X^2]$ . So:

$$\begin{aligned}
\mathbb{E}[X^2] &= \int_0^{\infty} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha+1} e^{-\beta x} dx \\
&= \frac{\beta^\alpha}{(\alpha)} \cdot \frac{\Gamma(\alpha + 2)}{\beta^{\alpha+2}} && (\text{Prop. 3}) \\
&= \frac{\beta^\alpha}{(\alpha)} \cdot \frac{(\alpha + 1)\Gamma(\alpha + 1)}{\beta^{\alpha+2}} && (\text{Prop. 2}) \\
&= \frac{\beta^\alpha}{(\alpha)} \cdot \frac{(\alpha + 1)(\alpha)\Gamma(\alpha)}{\beta^{\alpha+2}} && (\text{Prop. 2}) \\
&= \frac{\alpha^2 + \alpha}{\beta^2}
\end{aligned}$$

Substitute into the variance decomposition formula:

$$\begin{aligned}
\text{Var}(X) &= \frac{\alpha^2 + \alpha}{\beta^2} - \frac{\alpha^2}{\beta^2} \\
&= \frac{\alpha}{\beta^2} \quad \blacksquare
\end{aligned}$$

## Chapter 3

# Multivariate Probability

### 3.1 Expected Value of a Log-Normal Distribution

Given that the transformation of a variable,  $Y \sim N(\mu, \sigma^2)$ , given by  $X = e^Y$  results in a log-normal distribution, find the expected value of  $X$ .

#### 3.1.1 Solution:

Note that  $\mathbb{E}[X] = \mathbb{E}[e^Y]$ . As a result, we can simply use the normal distribution instead of a log-normal distribution:

$$\begin{aligned}\mathbb{E}[X] &= \mathbb{E}[e^Y] \\ &= \int_{-\infty}^{\infty} \frac{e^y}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}} dy \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2} + y} dy \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{y^2 + 2y\mu - \mu^2 + 2y\sigma^2}{2\sigma^2}} dy \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y^2 - 2(\mu + \sigma^2)y + \mu^2 + 2\sigma^2\mu + \sigma^4)}{2\sigma^2}} dy\end{aligned}$$

So far this is just algebra. We now complete the square inside the exponential. We want the end result to look like:  $(y - (\mu + \sigma^2))^2$ . To get this, we need to know what we should add to our expression. Multiplying out gives us:  $y^2 - 2(\mu + \sigma^2)y + \mu^2 + 2\sigma^2\mu + \sigma^4$ . Therefore we must multiply by  $e^{\frac{2\sigma^2\mu + \sigma^4 - 2\sigma^2\mu - \sigma^4}{2\sigma^2}}$ . This gives us:

$$\begin{aligned}
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(y-(\mu+\sigma^2))^2+2\sigma^2\mu+\sigma^4}{2\sigma^2}} dy \\
&= e^{\mu+\frac{1}{2}\sigma^2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(y-(\mu+\sigma^2))^2}{2\sigma^2}} dy
\end{aligned}$$

But note that the part leftover inside the integral is just a normal pdf! We know that this integrates to 1, which leaves us with:

$$= e^{\mu+\frac{1}{2}\sigma^2} \quad \blacksquare$$

### 3.2 Practice Problem 1: Hansen 4.1

Let  $f(x, y) = 1/4$  for  $-1 \leq x \leq 1$  and  $-1 \leq y \leq 1$ .

- Verify that  $f(x, y)$  is a valid density function
- Find the marginal density of  $X$
- Find the conditional density of  $Y$  given  $X = x$
- Find  $\mathbb{E}[Y|X = x]$
- Determine  $\mathbb{P}(X^2 + Y^2 \leq 1)$
- Determine  $\mathbb{P}(|X + Y| \leq 2)$

#### 3.2.1 Solution: Part a

We need to show that the joint pdf integrates to 1. So:

$$\begin{aligned}
\int_{-1}^1 \int_{-1}^1 \frac{1}{4} dy dx &= \int_{-1}^1 \left[ \frac{1}{4} \right]_{-1}^1 dx \\
&= \int_{-1}^1 \frac{1}{2} dx \\
&= \left[ \frac{1}{2}x \right]_{-1}^1 \\
&= \frac{1}{2} + \frac{1}{2} \\
&= 1 \quad \blacksquare
\end{aligned}$$

**3.2.2 Solution: Part b**

To find the marginal distribution for X, we simply integrate out Y.

$$\begin{aligned} f_x(x) &= \int_{-1}^1 \frac{1}{4} dy \\ &= \left[ \frac{1}{4} y \right]_{-1}^1 \\ &= \frac{1}{4} + \frac{1}{4} \\ &= \frac{1}{2} \quad \blacksquare \end{aligned}$$

Note that due to symmetry,  $f_y(y)$  is the same as  $f_x(x)$ .

**3.2.3 Solution: Part c**

We use the conditional density formula:  $f_{y|x}(y|x) = \frac{f(x,y)}{f(x)}$

$$\begin{aligned} f_{y|x}(y|x) &= \frac{1/4}{1/2} \\ &= \frac{1}{2} \quad \blacksquare \end{aligned}$$

**3.2.4 Solution: Part d**

Recall that  $\mathbb{E}[Y|X = x] = \int_{\mathbb{Y}} y f(y|x) dy$ . Applying this formula:

$$\begin{aligned} \mathbb{E}[Y|X = x] &= \int_{-1}^1 \frac{1}{2} y dy \\ &= \left[ \frac{1}{4} y^2 \right]_{-1}^1 \\ &= \frac{1}{4} - \frac{1}{4} \\ &= 0 \quad \blacksquare \end{aligned}$$

This is the same as  $\mathbb{E}[Y]$ . Why? Because  $f_{y|x}(y|x) = f_y(y)$ , meaning that  $X \perp Y$ .

**3.2.5 Solution: Part e**

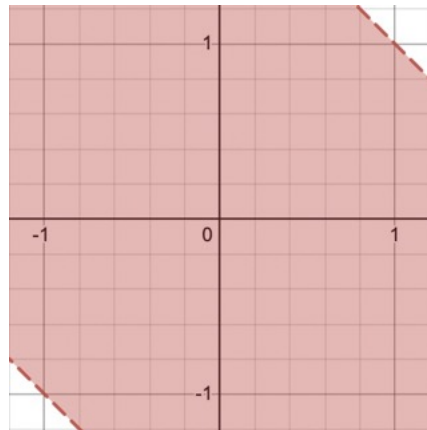
There are two ways to do this problem. I will use integrals for (e) and solve (f) graphically.

$$\begin{aligned}
\mathbb{P}(x^2 + y^2 \leq 1) &= \mathbb{P}(x^2 \leq 1 - y^2) \\
&= \mathbb{P}(-\sqrt{1 - y^2} \leq x \leq \sqrt{1 - y^2}) \\
&= \int_{-1}^1 \int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} \frac{1}{4} dx dy \\
&= \int_{-1}^1 \frac{\sqrt{1-y^2}}{2} dy && \text{using an integral calculator} \\
&= \left[ \frac{\arcsin(y) + y\sqrt{1-y^2}}{4} \right]_{-1}^1 \\
&= \frac{\pi}{4} \quad \blacksquare
\end{aligned}$$

Graphically, we can calculate the area of the square from the support of  $X$  and  $Y$  and the area of the circle enclosed by  $X^2 + Y^2 \leq 1$ . Doing this yields an area of 4 for the square and an area of  $\pi$  for the circle. Normalizing the space to an area of 1 by dividing by 4 leaves us with a probability equal to  $\frac{\pi}{4}$  confirming our integration solution.

### 3.2.6 Solution: Part f

As stated above, we will solve this graphically as solving via integration is tricky. Drawing a picture reveals the simplicity of this problem:



Where the shaded red area is every combination of  $X$  and  $Y$  such that their sum is less than two. Since our domain for each variable is between -1 and 1,  $\mathbb{P}(|X + Y| \leq 2) = 1$ .

### 3.3 Practice Problem 2: Hansen 4.7

Let  $X$  and  $Y$  have joint density  $f(x, y) = e^{-x-y}$  for  $x > 0$  and  $y > 0$ . Find  $f_x(x)$  and  $f_y(y)$  to determine if  $X \perp Y$ .

#### 3.3.1 Solution:

To find the marginal densities, we integrate out the other variable. So:

$$\begin{aligned} f_x(x) &= \int_0^{\infty} e^{-x-y} dy \\ &= [-e^{-x-y}]_0^{\infty} \\ &= 0 + e^{-x} \\ &= e^{-x} \end{aligned}$$

Similarly:

$$\begin{aligned} f_y(y) &= \int_0^{\infty} e^{-x-y} dx \\ &= [-e^{-x-y}]_0^{\infty} \\ &= 0 + e^{-y} \\ &= e^{-y} \end{aligned}$$

Recall that  $X$  and  $Y$  are  $\perp$  if  $f_x(x) \cdot f_y(y) = f(x, y)$ .

$$\begin{aligned} f_x(x) \cdot f_y(y) &= e^{-x} \cdot e^{-y} \\ &= e^{-x-y} \\ &= f(x, y) \quad \blacksquare \end{aligned}$$

So  $X$  and  $Y$  are  $\perp$ .

### 3.4 Practice Problem 3: Multivariate Change of Variables

Using the joint pdf from the previous problem ( $f(x, y) = e^{-x-y}$ ), apply the two transformations  $Z = X - Y$  and  $W = X + Y$ . Find the marginal distributions for  $Z$  and  $W$ .<sup>1</sup>

---

<sup>1</sup>Credit to [Penn State](#)

### 3.4.1 Solution: Finding the inverse functions

We begin with solving for  $X$ . Rearranging the transformation equations, we find that  $Z + Y = X$  and  $W - Y = X$ . Setting these equal:

$$\begin{aligned} Z + Y &= W - Y \\ 2Y &= W - Z \\ Y &= \frac{W - Z}{2} \end{aligned}$$

Similarly, by solving for  $Y$ , we find that:

$$\begin{aligned} -Z + X &= W - X \\ 2X &= W + Z \\ X &= \frac{W + Z}{2} \end{aligned}$$

### 3.4.2 Solution: Plug inverse functions into the joint pdf

$$\begin{aligned} f_{xy}(x, y) &= e^{-x-y} \\ f(g^{-1}(z, w), h^{-1}(z, w)) &= e^{-\frac{w-z}{2} + \frac{z-w}{2}} \\ &= e^{-\frac{2w}{2}} \\ &= e^{-w} \end{aligned}$$

### 3.4.3 Solution: The Jacobian

The Jacobian is the same as always:

$$\begin{aligned} |J| &= \begin{vmatrix} \frac{\partial x}{\partial z} & \frac{\partial x}{\partial w} \\ \frac{\partial y}{\partial z} & \frac{\partial y}{\partial w} \end{vmatrix} \\ &= \begin{vmatrix} \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{vmatrix} \\ &= \frac{1}{4} + \frac{1}{4} \\ &= \frac{1}{2} \end{aligned}$$

### 3.4.4 Solution: Substitute into the change of variables formula

Using the same formula as before



$$\begin{aligned} f_{zw}(z, w) &= e^{-w} \cdot \frac{1}{2} \\ &= \frac{e^{-w}}{2} \end{aligned}$$

Now we have our joint pdf!

### 3.4.5 Solution: Finding the marginal pdfs

First we need to figure out the support for W and Z. We know that:

$$\begin{array}{ll} 0 < x < \infty & 0 < y < \infty \\ 0 < \frac{w+z}{2} < \infty & 0 < \frac{w-z}{2} < \infty \\ 0 < w+z < \infty & 0 < w-z < \infty \end{array}$$

From here we can determine that

$$-w < z < w \quad 0 < w < \infty$$

We are ready to find the marginal distribution for W:

$$\begin{aligned} f_w(w) &= \int_{-w}^w \frac{e^{-w}}{2} dz \\ &= \left[ \frac{1}{2} e^{-w} z \right]_{-w}^w \\ &= \frac{1}{2} e^{-w} w + \frac{1}{2} e^{-w} w \\ &= w e^{-w} \quad \text{for } 0 < w < \infty \end{aligned}$$

One more to go! We need to re-evaluate the support before we can continue. We can write the support as:

$$-z < w < \infty \quad z < w < \infty \quad \text{with } -\infty < z < \infty$$

So:

$$\begin{aligned}
 f_z(z) &= \begin{cases} \int_{-z}^{\infty} \frac{e^{-w}}{2} dw & \text{for } -z < w < \infty \\ \int_z^{\infty} \frac{e^{-w}}{2} dw & \text{for } z < w < \infty \end{cases} \\
 &= \begin{cases} \left[ \frac{-1}{2} e^{-w} \right]_{-z}^{\infty} & \text{for } z < 0 \\ \left[ \frac{-1}{2} e^{-w} \right]_z^{\infty} & \text{for } z > 0 \end{cases} \\
 &= \begin{cases} \frac{1}{2} e^z & \text{for } z < 0 \\ \frac{1}{2} e^{-z} & \text{for } z > 0 \end{cases}
 \end{aligned}$$

If we combine these two cases, we arrive at our answer:

$$f_z(z) = \frac{1}{2} e^{-|z|}$$

This is the Laplace distribution's pdf with a location parameter of zero and a scale parameter of 1.

## Chapter 4

# Dependent Random Variables

### 4.1 Previous Problem: Hansen 4.14

Let  $X_1 \sim \Gamma(r, 1)$  and  $X_2 \sim \Gamma(s, 1)$  be independent. Find the distribution of  $Y = X_1 + X_2$ .

#### 4.1.1 Solution:

Many of you did change of variables on the last homework. That approach works well and should be your first instinct. Using change of variables here, though, is not particularly elegant. A more efficient method of arriving at the answer is via the moment generating function of the gamma random variable. We begin by finding that:

$$\begin{aligned} M(t) &= \mathbb{E}[e^{tx}] && \text{where } X \sim \Gamma(\alpha, \beta) \\ &= \int_0^\infty e^{tx} \frac{1}{\Gamma(\alpha)} \beta^{-\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}} dx \\ &= \frac{1}{\Gamma(\alpha)} \beta^{-\alpha} \int_0^\infty x^{\alpha-1} e^{-\frac{x}{\beta} + tx} dx \\ &= \frac{1}{\Gamma(\alpha)} \beta^{-\alpha} \int_0^\infty x^{\alpha-1} e^{-[\frac{1}{\beta} - t]x} dx \end{aligned}$$

Note that we can use the gamma function's properties to evaluate the inside of the integral. Using property 3 in Hansen:

$$\begin{aligned} &= \frac{1}{\Gamma(\alpha)} \beta^{-\alpha} \cdot \Gamma(\alpha) \left[ \frac{\beta}{1 - \beta t} \right]^\alpha \\ &= \left( \frac{1}{1 - \beta t} \right)^\alpha \end{aligned}$$

Okay, now that we know the MGF for a gamma random variable (which we could just look up on the web), we can begin to solve. We start with finding the MGF for  $Y$ :

$$\begin{aligned}
\mathbb{E}[e^{ty}] &= \mathbb{E}[e^{t(x_1+x_2)}] \\
&= \mathbb{E}[e^{tx_1+tx_2}] \\
&= \mathbb{E}[e^{tx_1}e^{tx_2}] && \text{Now, due to independence:} \\
&= \mathbb{E}[e^{tx_1}]\mathbb{E}[e^{tx_2}] \\
&= \left(\frac{1}{1-t}\right)^r \cdot \left(\frac{1}{1-t}\right)^s \\
&= \left(\frac{1}{1-t}\right)^{r+s}
\end{aligned}$$

But this should look familiar! After all, it is just the MGF for a gamma random variable with a shape parameter of  $r + s$ . Therefore  $Y \sim \Gamma(r + s, 1)$

## 4.2 Practice Problem 1: Hansen 4.19

Consider the hierarchical distribution

$$\begin{aligned}
X|N &\sim \chi_{2N}^2 \\
N &\sim \text{Poisson}(\lambda)
\end{aligned}$$

Find

- (a)  $\mathbb{E}[X]$
- (b)  $\text{Var}(X)$

### 4.2.1 Solution: Part a

We use the law of iterated expectations:  $\mathbb{E}[X] = \mathbb{E}_Y \mathbb{E}[X|Y]$

$$\begin{aligned}
\mathbb{E}[X] &= \mathbb{E}_N \mathbb{E}[X|N] \\
&= \mathbb{E}[2N] \\
&= 2\lambda
\end{aligned}$$

### 4.2.2 Solution: Part b

We use the law of total variance:  $\text{Var}(X) = \mathbb{E}[\text{Var}(X|Y)] + \text{Var}(\mathbb{E}[X|Y])$

$$\begin{aligned}
\text{Var}(X) &= \mathbb{E}[\text{Var}(X|N)] + \text{Var}(\mathbb{E}[X|N]) \\
&= \mathbb{E}[4N] + \text{Var}(2N) \\
&= 4\lambda + 4\text{Var}(N) \\
&= 4\lambda + 4\lambda \\
&= 8\lambda
\end{aligned}$$

### 4.3 Practice Problem 2: AR(2) Process

An AR(2) process relates a random variable located in time period  $t$  to two lags of that random variable. We usually write this as:  $y_t = d + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t$  where  $\varepsilon_t \sim WN(0, \sigma^2)$ . Recall that white noise implies that  $\varepsilon_t$  has no autocorrelation but does not imply that  $\varepsilon_t$  is not independent from its past.

- (a) Find  $\mathbb{E}[y_t]$
- (b) Find  $\text{Var}(y_t)$
- (c) Find  $\gamma_0$ ,  $\gamma_1$ , and  $\gamma_2$
- (d) Find the impulse responses for a shock  $\varepsilon_t$  for  $k \in 0, 1, 2, 3, 4$
- (e) Find  $\mathbb{E}[y_{t+3}|t]$

#### 4.3.1 Solution: Part a

We just take the expectation of the AR(2) equation:

$$\begin{aligned}
\mathbb{E}[y_t] &= \mathbb{E}[d + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t] \\
\mu_y &= d + \phi_1 \mathbb{E}[y_{t-1}] + \phi_2 \mathbb{E}[y_{t-2}] + 0 \\
\mu_y &= d + (\phi_1 + \phi_2) \mu_y \\
\mu_y(1 - \phi_1 - \phi_2) &= d \\
\mu_y &= \frac{d}{1 - \phi_1 - \phi_2}
\end{aligned}$$

#### 4.3.2 Solution: Part b

Similarly, we take the variance of the AR(2) equation:

$$\begin{aligned} \text{Var}(y_t) &= \text{Var}(d + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t) \\ \gamma_0 &= \phi_1^2 \text{Var}(y_{t-1}) + \phi_2^2 \text{Var}(y_{t-2}) + 2\phi_1\phi_2 \text{Cov}(y_{t-1}, y_{t-2}) + \sigma^2 \\ \gamma_0 &= (\phi_1^2 + \phi_2^2)\gamma_0 + 2\phi_1\phi_2\gamma_1 + \sigma^2 \end{aligned}$$

### 4.3.3 Solution: Part c

In class, Drew proved that the autocorrelation function for an AR(2) is given by:  $\rho_j = \phi_1\rho_{j-1} + \phi_2\rho_{j-2}$ . Using this equation we can find what we need. First, let's find  $\gamma_1$ :

$$\begin{aligned} \gamma_1 &= \phi_1\gamma_0 + \phi_2\gamma_{-1} && \text{Recall that } \gamma_{-i} = \gamma_i \\ \gamma_1 &= \phi_1\gamma_0 + \phi_2\gamma_1 \\ \gamma_1 &= \frac{\phi_1\gamma_0}{1 - \phi_2} \end{aligned}$$

Plugging this into the equation we found for  $\gamma_0$  in part a:

$$\begin{aligned} \gamma_0 &= (\phi_1^2 + \phi_2^2)\gamma_0 + 2\phi_1\phi_2 \frac{\phi_1\gamma_0}{1 - \phi_2} + \sigma^2 \\ \gamma_0(1 - \phi_1^2 - \phi_2^2 - 2\phi_1\phi_2 \frac{\phi_1}{1 - \phi_2}) &= \sigma^2 \\ \gamma_0 &= \frac{\sigma^2}{1 - \phi_1^2 - \phi_2^2 - 2\phi_1\phi_2 \frac{\phi_1}{1 - \phi_2}} \end{aligned}$$

Plug this into the expression for  $\gamma_1$ :

$$\begin{aligned} \gamma_1 &= \frac{\phi_1\gamma_0}{1 - \phi_2} \\ \gamma_1 &= \frac{\phi_1}{1 - \phi_2} \left[ \frac{\sigma^2}{1 - \phi_1^2 - \phi_2^2 - 2\phi_1\phi_2 \frac{\phi_1}{1 - \phi_2}} \right] \end{aligned}$$

Lastly, use the ACF to find  $\gamma_2$ :

$$\begin{aligned} \gamma_2 &= \phi_1\gamma_1 + \phi_2\gamma_0 \\ \gamma_2 &= \frac{\phi_1^2}{1 - \phi_2} \left[ \frac{\sigma^2}{1 - \phi_1^2 - \phi_2^2 - 2\phi_1\phi_2 \frac{\phi_1}{1 - \phi_2}} \right] + \frac{\phi_2\sigma^2}{1 - \phi_1^2 - \phi_2^2 - 2\phi_1\phi_2 \frac{\phi_1}{1 - \phi_2}} \end{aligned}$$

What if we did not have the ACF provided? We could always calculate  $\gamma_0$ ,  $\gamma_1$ , and  $\gamma_2$  by brute force. First, we rewrite the AR(2):

$$\begin{aligned}
y_t &= d + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t \\
y_t &= \mu_y(1 - \phi_1 - \phi_2) + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t \\
(y_t - \mu_y) &= \phi_1(y_{t-1} - \mu_y) + \phi_2(y_{t-2} - \mu_y) + \varepsilon_t
\end{aligned}$$

If we wanted to calculate the unconditional variance  $\gamma_0$ , for example, then:

$$\begin{aligned}
(y_t - \mu_y)^2 &= [\phi_1(y_{t-1} - \mu_y) + \phi_2(y_{t-2} - \mu_y) + \varepsilon_t](y_t - \mu_y) \\
\mathbb{E}[(y_t - \mu_y)^2] &= \mathbb{E}[(\phi_1(y_{t-1} - \mu_y) + \phi_2(y_{t-2} - \mu_y) + \varepsilon_t)(y_t - \mu_y)] \\
&= \gamma_0
\end{aligned}$$

What if we wanted to calculate  $\gamma_1$ ?

$$\begin{aligned}
(y_t - \mu_y)(y_{t-1} - \mu_y) &= [\phi_1(y_{t-1} - \mu_y) + \phi_2(y_{t-2} - \mu_y) + \varepsilon_t](y_{t-1} - \mu_y) \\
\mathbb{E}[(y_t - \mu_y)(y_{t-1} - \mu_y)] &= \mathbb{E}[(\phi_1(y_{t-1} - \mu_y) + \phi_2(y_{t-2} - \mu_y) + \varepsilon_t)(y_{t-1} - \mu_y)] \\
&= \gamma_1
\end{aligned}$$

Similarly for  $\gamma_2$ :

$$\begin{aligned}
(y_t - \mu_y)(y_{t-2} - \mu_y) &= [\phi_1(y_{t-1} - \mu_y) + \phi_2(y_{t-2} - \mu_y) + \varepsilon_t](y_{t-2} - \mu_y) \\
\mathbb{E}[(y_t - \mu_y)(y_{t-2} - \mu_y)] &= \mathbb{E}[(\phi_1(y_{t-1} - \mu_y) + \phi_2(y_{t-2} - \mu_y) + \varepsilon_t)(y_{t-2} - \mu_y)] \\
&= \gamma_2
\end{aligned}$$

#### 4.3.4 Solution: Part d

Impulses are defined as  $\frac{\partial y_{t+k}}{\partial \varepsilon_t}$ . Let's find this for  $k = 0$  first:

$$\begin{aligned}
y_t &= d + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t \\
\frac{\partial y_t}{\partial \varepsilon_t} &= 1
\end{aligned}$$

Now for  $k = 1$ :

$$\begin{aligned}
y_{t+1} &= d + \phi_1 y_t + \phi_2 y_{t-1} + \varepsilon_{t+1} \\
y_{t+1} &= d + \phi_1 (d + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t) + \phi_2 y_{t-1} + \varepsilon_{t+1} \\
\frac{\partial y_{t+1}}{\partial \varepsilon_t} &= \phi_1
\end{aligned}$$

For  $k = 2$ :

$$\begin{aligned}
y_{t+2} &= d + \phi_1 y_{t+1} + \phi_2 y_t + \varepsilon_{t+2} \\
\frac{\partial y_{t+2}}{\partial \varepsilon_t} &= \phi_1 \cdot \frac{\partial}{\partial \varepsilon_t} y_{t+1} + \phi_2 \cdot \frac{\partial}{\partial \varepsilon_t} y_t \\
&= \phi_1^2 + \phi_2
\end{aligned}$$

And for  $k = 3$ :

$$\begin{aligned}
y_{t+3} &= d + \phi_1 y_{t+2} + \phi_2 y_{t+1} + \varepsilon_{t+3} \\
\frac{\partial y_{t+3}}{\partial \varepsilon_t} &= \phi_1 \cdot \frac{\partial}{\partial \varepsilon_t} y_{t+2} + \phi_2 \cdot \frac{\partial}{\partial \varepsilon_t} y_{t+1} \\
&= \phi_1^3 + 2\phi_1\phi_2
\end{aligned}$$

Finally for  $k = 4$ :

$$\begin{aligned}
y_{t+4} &= d + \phi_1 y_{t+3} + \phi_2 y_{t+2} + \varepsilon_{t+4} \\
\frac{\partial y_{t+4}}{\partial \varepsilon_t} &= \phi_1 \cdot \frac{\partial}{\partial \varepsilon_t} y_{t+3} + \phi_2 \cdot \frac{\partial}{\partial \varepsilon_t} y_{t+2} \\
&= \phi_1^4 + 3\phi_1^2\phi_2 + \phi_2^2
\end{aligned}$$

As you can see, calculating impulse responses for an AR(2) by hand is much more labor-intensive than calculating them for an AR(1).

### 4.3.5 Solution: Part e

This is similar to calculating the unconditional moment. Now, though, we know all the variables up to time  $t$ . So:



$$\begin{aligned}
\mathbb{E}[y_{t+3}|t] &= \mathbb{E}[d + \phi_1 y_{t+2} + \phi_2 y_{t+1} + \varepsilon_{t+3}|t] \\
&= d + \phi_1 \mathbb{E}[y_{t+2}|t] + \phi_2 \mathbb{E}[y_{t+1}|t] + 0 \\
&= d + \phi_1 \mathbb{E}[d + \phi_1 y_{t+1} + \phi_2 y_t + \varepsilon_{t+2}|t] + \phi_2 \mathbb{E}[d + \phi_1 y_t + \phi_2 y_{t-1} + \varepsilon_{t+1}|t] \\
&= d + \phi_1 (d + \phi_2 y_t + \phi_1 \mathbb{E}[d + \phi_1 y_t + \phi_2 y_{t-1} + \varepsilon_{t+1}|t]) + \phi_2 (d + \phi_1 y_t + \phi_2 y_{t-1}) \\
&= d + \phi_1 d + \phi_1 \phi_2 y_t + \phi_1^2 d + \phi_1^3 y_t + \phi_1^2 \phi_2 y_{t-1} + \phi_2^2 y_{t-1} + \phi_2 d + \phi_2 \phi_1 y_t + \phi_2^2 y_{t-1} \\
&= d(1 + \phi_1 + \phi_1^2 + \phi_2) + y_t(\phi_1^3 + 2\phi_1 \phi_2) + y_{t-1}(\phi_1^2 \phi_2 + \phi_2^2)
\end{aligned}$$

Notice how we treated any variable dated at time  $t$  or before as if they were constants. That's the key to solving conditional expectations or variances with respect to time.

## 4.4 Practice Problem 3: Deriving the MA( $\infty$ ) Form for an AR(1)

Drew went over this in class, but this is an important wrench to have in your toolbox. The MA( $\infty$ ) form splits your autoregressive process into the stationary mean and sum of impulse shocks. We start with a basic AR(1):

$$\begin{aligned}
y_t &= d + \phi y_{t-1} + \varepsilon_t && \text{Recursively sub-in:} \\
&= d + \phi(d + \phi y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\
&= d(1 + \phi) + \phi^2 y_{t-2} + \phi \varepsilon_{t-1} + \varepsilon_t \\
&= d(1 + \phi) + \phi^2(d + \phi y_{t-3} + \varepsilon_{t-2}) + \phi \varepsilon_{t-1} + \varepsilon_t \\
&= d(1 + \phi + \phi^2) + \phi^3 y_{t-3} + \phi^2 \varepsilon_{t-2} + \phi \varepsilon_{t-1} + \varepsilon_t \\
&= d(1 + \phi + \phi^2) + \phi^3(d + \phi y_{t-4} + \varepsilon_{t-3}) + \phi^2 \varepsilon_{t-2} + \phi \varepsilon_{t-1} + \varepsilon_t \\
&= d(1 + \phi + \phi^2 + \phi^3) + \phi^4 y_{t-4} + \phi^3 \varepsilon_{t-3} + \phi^2 \varepsilon_{t-2} + \phi \varepsilon_{t-1} + \varepsilon_t
\end{aligned}$$

We see a pattern emerging here. Using induction, we can write this process as:

$$= d \sum_{i=1}^{\infty} \phi^{i-1} + \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}$$

Using the fact that  $\sum_{k=0}^{\infty} ar^k = \frac{a}{1-r}$ , and making the assumption that  $\phi < 1$ , this becomes:

$$= \frac{d}{1-\phi} + \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}$$

We can simplify further. Solving for the mean of an AR(1) process:

$$\begin{aligned}\mathbb{E}[y_t] &= \mathbb{E}[d + \phi y_{t-1} + \varepsilon_t] \\ \mu_y &= d + \phi \mathbb{E}[y_{t-1}] + 0 \\ \mu_y &= d + \phi \mu_y \\ \mu_y &= \frac{d}{1-\phi}\end{aligned}$$

Therefore, the MA( $\infty$ ) form becomes:

$$y_t = \mu_y + \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}$$

Where  $\mu_y$  is the unconditional mean of  $Y$  and the sum contains the impulse shocks of  $\varepsilon$ .

## Chapter 5

# Dependent Vectors of Random Variables

### 5.1 Previous Problem 1: Hansen 5.12

Show that the MGF of  $X \sim N(\mu, \Sigma) \in \mathbb{R}^m$  is

$$\begin{aligned} M(t) &= \mathbb{E}[e^{t^T X}] \\ &= e^{t^T \mu + \frac{1}{2} t^T \Sigma t} \end{aligned}$$

#### 5.1.1 Solution:

We start from the definition of the moment generating function:

$$\begin{aligned} M_x(t) &= \mathbb{E}[e^{tx}] \\ &= \frac{1}{\sqrt{2\pi}} \left| \Sigma^{-1/2} \right| \int_{-\infty}^{\infty} e^{t^T x} \cdot e^{-\frac{1}{2} (x-\mu)^T \Sigma^{-1} (x-\mu)} dx \\ &= \frac{1}{\sqrt{2\pi}} \left| \Sigma^{-1/2} \right| \int_{-\infty}^{\infty} e^{-\frac{1}{2} (x^T \Sigma^{-1} x - \mu^T \Sigma^{-1} x - x^T \Sigma^{-1} \mu + \mu^T \Sigma^{-1} \mu) + t^T x} dx \\ &= \frac{1}{\sqrt{2\pi}} \left| \Sigma^{-1/2} \right| \int_{-\infty}^{\infty} e^{-\frac{1}{2} (x^T \Sigma^{-1} x - \mu^T \Sigma^{-1} x - x^T \Sigma^{-1} \mu + \mu^T \Sigma^{-1} \mu - 2t^T x)} dx \\ &= \frac{1}{\sqrt{2\pi}} \left| \Sigma^{-1/2} \right| e^{-\frac{\mu^T \Sigma^{-1} \mu}{2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2} (x^T \Sigma^{-1} x - (\mu^T + t \Sigma) \Sigma^{-1} x - x^T \Sigma^{-1} \mu - t^T x)} dx \\ &= \frac{1}{\sqrt{2\pi}} \left| \Sigma^{-1/2} \right| e^{-\frac{\mu^T \Sigma^{-1} \mu}{2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2} (x^T \Sigma^{-1} x - (\mu^T + t \Sigma) \Sigma^{-1} x - (\mu^T \Sigma^{-1} x)^T - t^T x)} dx \end{aligned}$$

$$= \frac{1}{\sqrt{2\pi}} \left| \Sigma^{-1/2} \right| e^{-\frac{\mu^T \Sigma^{-1} \mu}{2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x^T \Sigma^{-1} x - (\mu^T + t \Sigma)^T \Sigma^{-1} x - (\mu^T + t \Sigma)^T \Sigma^{-1} x)} dx$$

Now we need to complete the square inside the integrand's exponential. We can see that we will want:  $(x^T - (\mu^T + t \Sigma)^T) \Sigma^{-1} (x - (\mu + t \Sigma))$ . Multiplying this out to see what we need:

$$\begin{aligned} (x^T - (\mu^T + t \Sigma)^T) \Sigma^{-1} (x - (\mu + t \Sigma)) &= x^T \Sigma^{-1} x - x^T \Sigma^{-1} (\mu + t \Sigma) - (\mu^T + t \Sigma)^T \Sigma^{-1} x \\ &\quad + (\mu^T + t \Sigma)^T \Sigma^{-1} (\mu + t \Sigma) \end{aligned}$$

So we need the last term. Therefore we multiply by:

$$e^{-\frac{1}{2}((\mu^T + t \Sigma)^T \Sigma^{-1} (\mu + t \Sigma) - (\mu^T + t \Sigma)^T \Sigma^{-1} (\mu + t \Sigma))}$$

Now we have:

$$\begin{aligned} &= \frac{1}{\sqrt{2\pi}} \left| \Sigma^{-1/2} \right| e^{-\frac{\mu^T \Sigma^{-1} \mu}{2}} e^{-\frac{1}{2}((\mu^T + t \Sigma)^T \Sigma^{-1} (\mu + t \Sigma) - (\mu^T + t \Sigma)^T \Sigma^{-1} (\mu + t \Sigma))} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x^T - (\mu^T + t \Sigma)^T) \Sigma^{-1} (x - (\mu + t \Sigma))} dx \\ &= e^{\frac{1}{2}(-\mu^T \Sigma^{-1} \mu + \mu^T \Sigma^{-1} \mu + 2\mu^T t + t^T t)} \\ &= e^{\mu^T t + \frac{1}{2} t^T \Sigma t} \quad \blacksquare \end{aligned}$$

## 5.2 Previous Problem 2: Deriving the F Distribution

Define the random variables  $U$  and  $V$  as:

$$\begin{aligned} U &\sim \chi^2(k) \\ V &\sim \chi^2(a) \end{aligned}$$

Assume that  $U$  and  $V$  are independent. Define the random variable  $F$  as:

$$F = \frac{U}{k} \cdot \frac{a}{V}$$

We will derive the F distribution and expectation.

### 5.2.1 Solution:

We begin with finding the joint distribution of  $U$  and  $V$ . Because they are independent, we can multiply their pdfs together. After doing so and combining like terms, we arrive at:

$$f(u, v) = \frac{1}{2^{\frac{k+a}{2}} \Gamma(\frac{k}{2}) \Gamma(\frac{a}{2})} u^{\frac{k}{2}-1} v^{\frac{a}{2}-1} e^{-\frac{(u+v)}{2}}$$

We now need to do a change of variables. Rearranging the definition of F:

$$F = \frac{U}{k} \cdot \frac{a}{V}$$

$$U = \frac{kFV}{a}$$

So we map  $U$  to  $\frac{kFV}{a}$  and  $V$  to  $V$ . The next step is finding the Jacobian:

$$|J| = \begin{vmatrix} \frac{\partial U}{\partial F} & \frac{\partial U}{\partial V} \\ \frac{\partial V}{\partial F} & \frac{\partial V}{\partial V} \end{vmatrix}$$

$$= \begin{vmatrix} \frac{kV}{a} & \frac{kF}{a} \\ 0 & 1 \end{vmatrix}$$

$$= \frac{kV}{a}$$

Plug our inverse function into the joint pdf and multiply by the Jacobian to get (after combining like terms):

$$f(F, V) = \frac{k}{a} \frac{1}{2^{\frac{k+a}{2}} \Gamma(\frac{k}{2}) \Gamma(\frac{a}{2})} \left( \frac{kFV}{a} \right)^{\frac{k}{2}-1} V^{\frac{a}{2}} e^{-\frac{(kFV}{a} + V)/2}$$

To find the pdf of F, we simply integrate out V:

$$f(F) = \int_{-\infty}^{\infty} \frac{k}{a} \frac{1}{2^{\frac{k+a}{2}} \Gamma(\frac{k}{2}) \Gamma(\frac{a}{2})} \left( \frac{kFV}{a} \right)^{\frac{k}{2}-1} V^{\frac{a}{2}} e^{-\frac{(kFV}{a} + V)/2} dV$$

$$= \frac{k}{a} \frac{1}{2^{\frac{k+a}{2}} \Gamma(\frac{k}{2}) \Gamma(\frac{a}{2})} \int_{-\infty}^{\infty} \left( \frac{kFV}{a} \right)^{\frac{k}{2}-1} V^{\frac{a}{2}} e^{-\frac{(kFV}{a} + V)/2} dV$$

$$= \frac{k}{a} \frac{1}{2^{\frac{k+a}{2}} \Gamma(\frac{k}{2}) \Gamma(\frac{a}{2})} \left( \frac{kF}{a} \right)^{\frac{k}{2}-1} \int_{-\infty}^{\infty} V^{\frac{a+k}{2}-1} e^{-\frac{(kFV}{a} + V)/2} dV$$

Note that this is the kernel of  $\Gamma\left(\frac{a+k}{2}, \frac{1}{2}\left(\frac{kF}{a} + 1\right)\right)$ . By multiplying and dividing by the appropriate normalizing constant we can integrate the integrand out to 1, leaving us with:

$$f(F) = \frac{k}{a} \frac{\Gamma\left(\frac{a+k}{2}\right)}{2^{\frac{k+a}{2}} \Gamma(\frac{k}{2}) \Gamma(\frac{a}{2})} \left( \frac{kF}{a} \right)^{\frac{k}{2}-1} \left( \frac{1}{2} \cdot \left( \frac{kF}{a} + 1 \right) \right)^{-\frac{(a+k)}{2}}$$

Which is the pdf for the F distribution. One last step! Now, we either calculate the expected value for the F distribution the easy way or the hard way. This problem misleads us into thinking that the easy way is to use the pdf we just derived. The far easier path though is as follows:

$$\begin{aligned}
 \mathbb{E}[F] &= \mathbb{E}\left[\frac{U}{k} \cdot \frac{a}{V}\right] \\
 &= \frac{a}{k} \mathbb{E}\left[\frac{U}{V}\right] && \text{Using the independence of } U \text{ and } V: \\
 &= \frac{a}{k} \mathbb{E}[U] \mathbb{E}\left[\frac{1}{V}\right] && \text{Using the mean of a } \chi^2 \text{ and inverse } \chi^2: \\
 &= \frac{a}{k} (k) \left(\frac{1}{a-2}\right) \\
 &= \frac{a}{a-2} \quad \blacksquare
 \end{aligned}$$

### 5.3 Practice Problem 1: Companion Form

Companion form basically rewrites any linear model (that has finite memory) into a model that looks like a VAR(1). As such, we can determine the mean, variance, covariance, and whether a linear model is stationary from this form. In general:

$$\boldsymbol{\alpha}_t = \mathbf{d} + \mathbf{T}\boldsymbol{\alpha}_{t-1} + \mathbf{R}\boldsymbol{\varepsilon}_t$$

is known as the companion form where  $\boldsymbol{\varepsilon}_t \sim W.N.(0, \mathbf{Q})$ . Structure-wise,  $\boldsymbol{\alpha}_t$  is  $mx1$ ,  $\mathbf{d}$  is  $mx1$ ,  $\mathbf{T}$  is  $mxm$ ,  $\mathbf{R}$  is  $mxq$ , and  $\mathbf{Q}$  is  $qxq$ .

By looking at how the companion form is written, we note that  $\mathbf{Q}$  is the variance-covariance matrix of the shocks,  $\mathbf{T}$  is the matrix placing weights on the right-hand side variables,  $\mathbf{d}$  is the vector of means, and  $\mathbf{R}$  is vector of weights on the shock term.  $\boldsymbol{\alpha}$  is the vector of variables in general, leaving out the last shock lag and/or last variable lag.

The easiest way to demonstrate how companion form can be used is by working through an example. Let's look at an ARMA(3,1) model. First, let's write the model as we are used to seeing it:

$$y_t = d + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \theta \varepsilon_{t-1} + \varepsilon_t$$

This should look familiar. Looking at our companion form rules we can begin to build our matrices. Starting with  $\boldsymbol{\alpha}_t$ :

$$\boldsymbol{\alpha}_t = \begin{bmatrix} y_t \\ y_{t-1} \\ y_{t-2} \\ \varepsilon_t \end{bmatrix}$$

Notice how I left out  $y_{t-3}$  and  $\varepsilon_{t-1}$ . Next, we build the  $\mathbf{d}$  vector:

$$\mathbf{d} = \begin{bmatrix} d \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

That's not too difficult to build. Now we go onto the  $\mathbf{T}$  matrix:

$$\mathbf{T} = \begin{bmatrix} \phi_1 & \phi_2 & \phi_3 & \theta \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

and then the  $\boldsymbol{\alpha}_{t-1}$  vector:

$$\boldsymbol{\alpha}_{t-1} = \begin{bmatrix} y_{t-1} \\ y_{t-2} \\ y_{t-3} \\ \varepsilon_{t-1} \end{bmatrix}$$

Notice that by multiplying  $\mathbf{T}$  with  $\boldsymbol{\alpha}_{t-1}$  we get the terms with the coefficients in the ARMA process. And then, because  $y_{t-1}$  must equal  $y_{t-1}$  and  $y_{t-2}$  must equal  $y_{t-2}$  we place ones in the appropriate places. Only one more vector to go:

$$\mathbf{R} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

Putting all of these together, we get the companion form:

$$\begin{bmatrix} y_t \\ y_{t-1} \\ y_{t-2} \\ \varepsilon_t \end{bmatrix} = \begin{bmatrix} d \\ 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} \phi_1 & \phi_2 & \phi_3 & \theta \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ y_{t-2} \\ y_{t-3} \\ \varepsilon_{t-1} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} \varepsilon_t$$

Companion forms are not unique. Rather, we can swap rows to create "new" companion forms. Each variation, if done correctly, should return the same forecasts and unconditional moments as the others.

I claimed earlier that companion form is useful for testing stationarity. How do we test this? We want to see if the coefficients on the middle terms lead to exploding  $y_t$  values. To do so, we can check if the eigenvalues of our slope matrix ( $\mathbf{T}$ ) are  $< 1$ .<sup>2</sup> Recall that to check the eigenvalues of a matrix, we use:

$$|\mathbf{T} - \lambda\mathbf{I}| = 0$$

If the eigenvalues ( $\lambda$ ) are  $\geq 1$  our process is not stationary.

I also claimed that we can easily find the unconditional moments. Conveniently, if  $\boldsymbol{\alpha}_t$  is stationary, the unconditional moments follow formulaically:

$$\begin{aligned} \mathbb{E}[\boldsymbol{\alpha}_t] &= (\mathbf{I} - \mathbf{T})^{-1}\mathbf{d} \\ &= \boldsymbol{\mu}_\alpha \end{aligned}$$

$$\begin{aligned} \text{vec}(\text{var}(\boldsymbol{\alpha}_t)) &= (\mathbf{I}_{m^2} - (\mathbf{T} \otimes \mathbf{T}))^{-1}\text{vec}(\mathbf{RQR}^T) \\ &= \text{vec}(\boldsymbol{\Gamma}_0) \end{aligned}$$

$$\begin{aligned} \text{cov}(\boldsymbol{\alpha}_t, \boldsymbol{\alpha}_{t-j}) &= \mathbf{T}^j \boldsymbol{\Gamma}_0 \\ &= \boldsymbol{\Gamma}_j \end{aligned}$$

Let's apply these formulae to our ARMA(3,1) process. First, we solve for the eigenvalues:

---

<sup>2</sup>Technically, this is a necessary but not sufficient condition. In time-series econometrics, we often treat it as sufficient. For more information, see [Carnegie Mellon](#).



$$\begin{aligned}
 |\mathbf{T} - \lambda\mathbf{I}| &= \left| \begin{bmatrix} \phi_1 & \phi_2 & \phi_3 & \theta \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} - \begin{bmatrix} \lambda & 0 & 0 & 0 \\ 0 & \lambda & 0 & 0 \\ 0 & 0 & \lambda & 0 \\ 0 & 0 & 0 & \lambda \end{bmatrix} \right| \\
 &= \left| \begin{bmatrix} \phi_1 - \lambda & \phi_2 & \phi_3 & \theta \\ 1 & -\lambda & 0 & 0 \\ 0 & 1 & -\lambda & 0 \\ 0 & 0 & 0 & -\lambda \end{bmatrix} \right|
 \end{aligned}$$

Because this is quite large, I use a solver to calculate the determinant and get:

$$0 = -\lambda(-\lambda(-\lambda(\phi_1 - \lambda) - \phi_2) + \phi_3)$$

If the values of the coefficients are such that each root of the equation is, in absolute terms, less than 1 (on the real or complex unit circle), then the process is stationary. Here, Wolfram Alpha returns only one eigenvalue: 0. So the process is stationary.

Now let's solve for the unconditional mean:

$$\begin{aligned}
 \mathbb{E}[\boldsymbol{\alpha}_t] &= (\mathbf{I} - \mathbf{T})^{-1}\mathbf{d} \\
 &= \left( \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} - \begin{pmatrix} \phi_1 & \phi_2 & \phi_3 & \theta \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \right)^{-1} \begin{bmatrix} d \\ 0 \\ 0 \\ 0 \end{bmatrix}
 \end{aligned}$$

$$\begin{aligned}
&= \left( \begin{bmatrix} 1 - \phi_1 & -\phi_2 & -\phi_3 & -\theta \\ -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \right)^{-1} \begin{bmatrix} d \\ 0 \\ 0 \\ 0 \end{bmatrix} && \text{Let } P \equiv 1 - \phi_1 - \phi_2 - \phi_3 \\
&= \frac{1}{P} \begin{bmatrix} 1 & \phi_3 + \phi_2 & \phi_3 & \theta \\ 1 & 1 - \phi_1 & \phi_3 & \theta \\ 1 & 1 - \phi_1 & 1 - \phi_1 - \phi_2 & \theta \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} d \\ 0 \\ 0 \\ 0 \end{bmatrix} \\
&= \begin{bmatrix} \frac{d}{P} \\ \frac{d}{P} \\ \frac{d}{P} \\ 0 \end{bmatrix}
\end{aligned}$$

This matches all of the results from previous problems! The mean is usually tractable to do by hand. The variance, though, due to the kronecker product can take up lots of space. A computer can easily handle the computation though (see Problem Set 5).

## 5.4 Matlab Help

This week's Matlab section is hard. Because it's hard, I'm going to provide you a little guidance so that you know that you're on the right track.

First, please use vectorization to generate random data. Do not use loops. The run time for your code can explode. Here's an example:

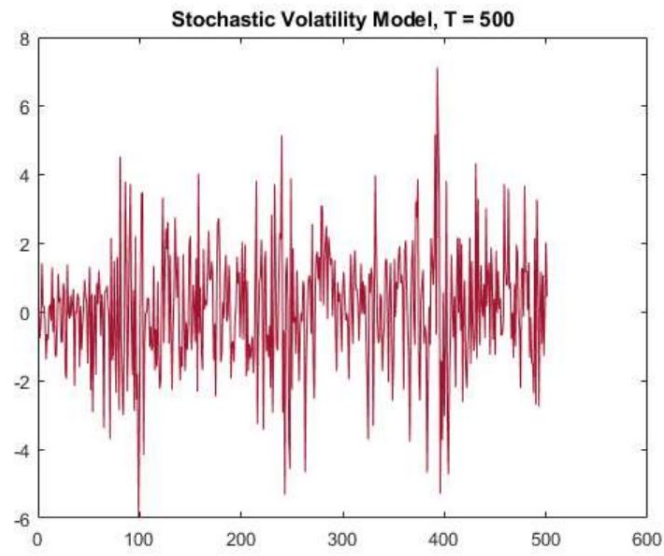
```

mu = 1;
phi = .97;
sigma = sqrt(.025);
e = normrnd(0,1, 5004,1);
n = normrnd(0,1,5001,1);

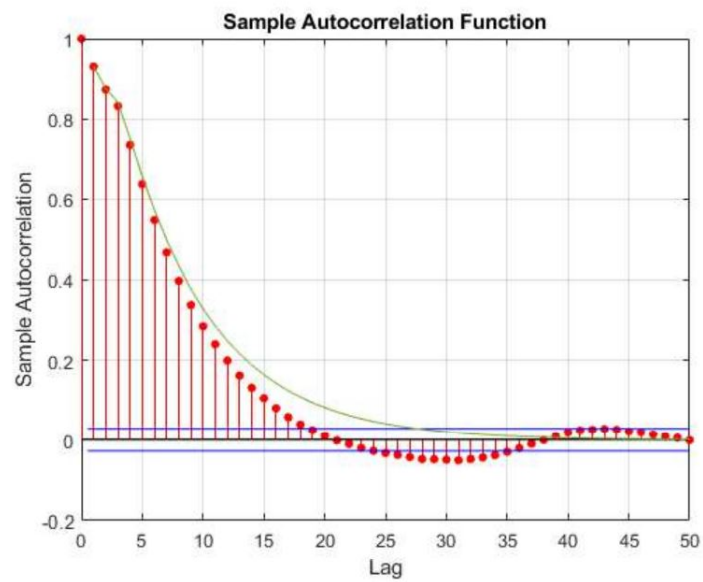
```

Notice how the variables "e" and "n" are built. Instead of having to loop, Matlab can vectorize the process.

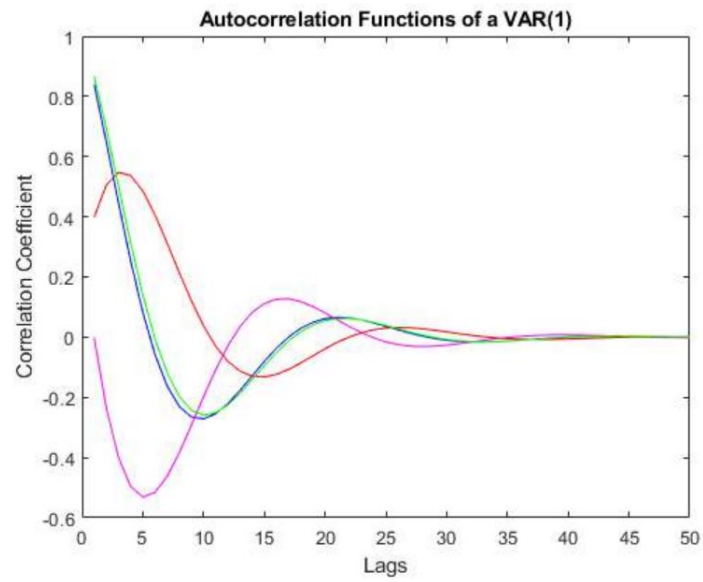
Here is how your graphs from question four should look:



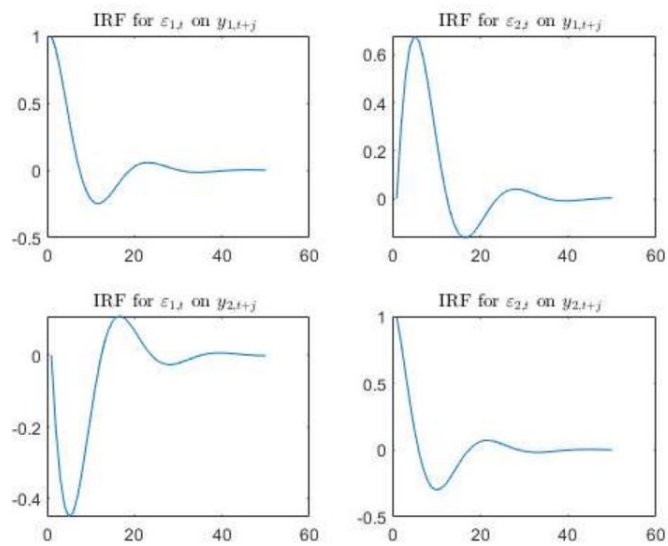
Here is how your sample ACF and theoretical ACF should approximately line up:



Here is how your VAR(1) ACFs should look:



And lastly, this is how your IRFs should look:



Good luck!

## Chapter 6

# Bias and Consistency

### 6.1 The Analogy Principle

Most of this is in Drew's slides, but this topic is important. I will cover the analogy principle quickly.

Suppose we have a function  $\beta = h(\theta)$  where  $\theta = \mathbb{E}[g(y_i)]$ . We want to estimate  $\beta$ . We do not know  $\mathbb{E}[g(y_i)]$ . Therefore, we replace  $\theta$  with  $\hat{\theta} = \frac{1}{n} \sum_{i=1}^n y_i$ . So  $\hat{\beta} = h(\hat{\theta})$ .

Let's go through a simple example using variance. Variance is given by :

$$\sigma^2 = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

In practice, replace the expectations with the sample mean. Therefore:

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \left( \frac{1}{n} \sum_{i=1}^n x_i \right)^2$$

We now have a viable estimate for variance.

### 6.2 Bias

The definition of bias is:

$$\mathbb{E}[\hat{\theta}] - \theta = \text{bias}$$

If  $\text{bias} = 0$ , we say that the estimator is unbiased. How does this apply to the analogy principle? Ideally, we'd like to create unbiased estimators for the parameters we are looking for. Let's look at the estimator for the variance that we just found.

First, we should check our two plug-ins to see if they are unbiased.

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n x_i^2\right] &= \frac{1}{n}\mathbb{E}[\sum_{i=1}^n x_i^2] \\
&= \frac{1}{n}\sum_{i=1}^n \mathbb{E}[x_i^2] \\
&= \frac{1}{n}\sum_{i=1}^n \mu'_2 \\
&= \frac{1}{n} \cdot n\mu'_2 \\
&= \mu'_2
\end{aligned}$$

So that part is unbiased. Let's check the standard mean:

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n x_i\right] &= \frac{1}{n}\mathbb{E}[\sum_{i=1}^n x_i] \\
&= \frac{1}{n}\sum_{i=1}^n \mathbb{E}[x_i] \\
&= \frac{1}{n}\sum_{i=1}^n \mu \\
&= \frac{1}{n} \cdot n\mu \\
&= \mu
\end{aligned}$$

So this is also unbiased. Let's check them together in the variance estimate:

$$\begin{aligned}
\mathbb{E}[\hat{\sigma}^2] &= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n x_i^2 - \left(\frac{1}{n}\sum_{i=1}^n x_i\right)^2\right] \\
&= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n x_i^2\right] - \mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^n x_i\right)^2\right] \\
&= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n x_i^2\right] - \mathbb{E}\left[\left(\mu - \mu + \frac{1}{n}\sum_{i=1}^n x_i\right)^2\right] \\
&= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n x_i^2\right] - \mathbb{E}\left[\left(\mu + \frac{1}{n}\sum_{i=1}^n (x_i - \mu)\right)^2\right] \\
&= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n x_i^2\right] - \mathbb{E}\left[\mu^2 + 2\mu\frac{1}{n}\sum_{i=1}^n (x_i - \mu) + \left(\frac{1}{n}\sum_{i=1}^n (x_i - \mu)\right)^2\right]
\end{aligned}$$

$$\begin{aligned}
&= \mathbb{E} \left[ \frac{1}{n} \sum_{i=1}^n x_i^2 \right] - \mathbb{E} [\mu^2 + 2\mu(\bar{x} - \mu) + (\bar{x} - \mu)^2] \\
&= \mu_2' - \mu^2 - 2\mu(\mu - \mu) - \mathbb{E}[(\bar{x} - \mu)^2] \\
&= \sigma^2 - \text{Var}(\bar{x}) \\
&= \sigma^2 - \text{Var} \left( \frac{1}{n} \sum_{i=1}^n x_i \right) \\
&= \sigma^2 - \frac{1}{n^2} \sum_{i=1}^n \text{Var}(x_i) \\
&= \sigma^2 - \frac{1}{n^2} \cdot n\sigma^2 \\
&= \sigma^2 - \frac{\sigma^2}{n} \\
&= \left( 1 - \frac{1}{n} \right) \sigma^2
\end{aligned}$$

Unfortunately, our estimator is biased, where the bias is  $-\frac{1}{n}\sigma^2$ . In class, Drew showed you that an unbiased estimator for variance is:

$$\hat{s}^2 = \frac{1}{n-1} \sum_{i=1}^n [y_i - \bar{y}]^2$$

To correct our estimator, we simply add  $\frac{\hat{s}^2}{n}$ :

$$\begin{aligned}
\hat{\sigma}_{unbiased}^2 &= \hat{\sigma}^2 + \frac{\hat{s}^2}{n} \\
\mathbb{E} [\hat{\sigma}_{unbiased}^2] &= \mathbb{E} \left[ \hat{\sigma}^2 + \frac{\hat{s}^2}{n} \right] \\
&= \mathbb{E}[\hat{\sigma}^2] + \mathbb{E} \left[ \frac{\hat{s}^2}{n} \right] \\
&= \sigma^2 - \frac{\sigma^2}{n} + \frac{\sigma^2}{n} \\
&= \sigma^2
\end{aligned}$$

Note that sometimes obtaining an unbiased estimator requires sacrificing precision. This trade-off gives rise to the bias-variance trade-off idea. One way to balance these two is to minimize the mean-squared error:

$$m.s.e(\hat{\theta}) = \text{Var}(\hat{\theta}) + (\text{bias}(\hat{\theta}))^2$$

## 6.3 Consistency

Let's complicate the issue now. When we take asymptotics, we are looking not at how the estimators behave in expectation, but how they behave as our sample size nears infinity. We have three theorems that will be helpful in this section.

The Continuous Mapping Theorem says that if  $x_n \xrightarrow{P} x$  then  $h(x_n) \xrightarrow{P} h(x)$ , for continuous  $h(\cdot)$ . The relationship also holds for convergence in distribution and convergence almost surely.

Slutsky's Theorem is an extension of the CMT. This theorem states that:

If  $x_n \xrightarrow{d} x$  and  $y_n \xrightarrow{P} y$ , then:

$$(i) \quad x_n + y_n \xrightarrow{d} x + y$$

$$(ii) \quad x_n y_n \xrightarrow{d} xy$$

$$(iii) \quad \frac{x_n}{y_n} \xrightarrow{d} \frac{x}{y}$$

Note the convergence requirements for  $x$  and  $y$ .

The last theorem is the Delta Method. You will use this theorem over and over again throughout the year with the Central Limit Theorem. Make sure you know it!

$$\sqrt{n}(h(x_n) - h(x)) \xrightarrow{d} N(0, H^T V H)$$

where  $H = \frac{\partial}{\partial u} h(u)^T$ .

### 6.3.1 Application to the Sample Variance

Recall our estimator for  $\sigma^2$  consisted of two parts:

$$\begin{aligned} \hat{\theta} &= \begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{n} \sum_{i=1}^n x_i \\ \frac{1}{n} \sum_{i=1}^n x_i^2 \end{bmatrix} \end{aligned}$$

By the Weak Law of Large Numbers  $\frac{1}{n} \sum_{i=1}^n x_i \xrightarrow{P} \mathbb{E}[x]$ . Similarly, the WLLN means that  $\frac{1}{n} \sum_{i=1}^n x_i^2 \xrightarrow{P} \mathbb{E}[x^2]$ . Recall the  $\sigma^2$  estimator:

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \left( \frac{1}{n} \sum_{i=1}^n x_i \right)^2$$

The difference of two continuous functions is continuous, so the CMT applies:



$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n x_i^2 - \left( \frac{1}{n} \sum_{i=1}^n x_i \right)^2 &\xrightarrow{P} \mathbb{E}[x^2] - \mathbb{E}[x]^2 \\ &= \sigma^2 \end{aligned}$$

What we have found is that while the sample variance is not unbiased, it is consistent. We now need to find the variance of our estimator using the Delta Method formula.

First, we start with  $V$  :

$$\begin{aligned} V &= \mathbb{E} \left[ \left( \begin{bmatrix} x \\ x^2 \end{bmatrix} - \begin{bmatrix} \mathbb{E}[x] \\ \mathbb{E}[x^2] \end{bmatrix} \right) \left( \begin{bmatrix} x \\ x^2 \end{bmatrix} - \begin{bmatrix} \mathbb{E}[x] \\ \mathbb{E}[x^2] \end{bmatrix} \right)^T \right] \\ &= \mathbb{E} \left[ \begin{bmatrix} x - \mathbb{E}[x] \\ x^2 - \mathbb{E}[x^2] \end{bmatrix} \begin{bmatrix} x - \mathbb{E}[x] \\ x^2 - \mathbb{E}[x^2] \end{bmatrix}^T \right] \end{aligned}$$

Note that  $h(\boldsymbol{\theta}) = \theta_2 - \theta_1^2$ . Now we find  $H$ :

$$\begin{aligned} H &= \frac{\partial}{\partial \boldsymbol{\theta}} h(\boldsymbol{\theta}) \\ &= \frac{\partial}{\partial \boldsymbol{\theta}} (\theta_2 - \theta_1^2) \\ &= \begin{bmatrix} 2\theta_1 \\ 1 \end{bmatrix} \end{aligned}$$

Putting this altogether, our asymptotic variance is:

$$H^T V H = \begin{bmatrix} 2\theta_1 \\ 1 \end{bmatrix} \mathbb{E} \left[ \begin{bmatrix} x - \mathbb{E}[x] \\ x^2 - \mathbb{E}[x^2] \end{bmatrix} \begin{bmatrix} x - \mathbb{E}[x] \\ x^2 - \mathbb{E}[x^2] \end{bmatrix}^T \right] \begin{bmatrix} 2\theta_1 \\ 1 \end{bmatrix}^T$$

Therefore, via the Delta Method:

$$\sqrt{n}(\hat{\sigma}^2 - \sigma^2) \xrightarrow{d} N \left( 0, \begin{bmatrix} 2\theta_1 \\ 1 \end{bmatrix} \mathbb{E} \left[ \begin{bmatrix} x - \mathbb{E}[x] \\ x^2 - \mathbb{E}[x^2] \end{bmatrix} \begin{bmatrix} x - \mathbb{E}[x] \\ x^2 - \mathbb{E}[x^2] \end{bmatrix}^T \right] \begin{bmatrix} 2\theta_1 \\ 1 \end{bmatrix}^T \right)$$

## 6.4 Midterm Review

We have covered an advanced probability course in less than a half-semester. The amount of material you have learned over the last month and a half is not insignificant. In an effort to help organize your studying for the midterm, I have created a list of the big topics.

### 6.4.1 Basics of Probability

- (a) Probability rules
  - (i) Conditional Probability
  - (ii) Bayes' Rule
  - (iii) Law of Total Probability
  - (iv) Monte Hall Problem
- (b) Properties of the CDF and PDF
- (c) Converting between CDFs and PDFs

### 6.4.2 Distributions

- (a) Joint, marginal, and conditional PDFs
- (b) Univariate and Multivariate Change of Variables
- (c) Calculating expectations, variances, covariances
- (d) Law of Total Variance and Law of Iterated Expectations
- (e) Distribution kernels and normalizing constants
- (f) Moment Generating Functions

### 6.4.3 Dependent Random Variables

- (a) Common representations
  - (i) MA(q)
  - (ii) AR(p)
  - (iii) ARMA(p,q)
  - (iv) MA( $\infty$ ) form
  - (v) VAR(p)
  - (vi) White Noise assumptions
- (b) Unconditional expectations, variances, covariances
  - (i) From standard form
  - (ii) From MA( $\infty$ ) form
  - (iii) From Companion form
- (c) Impulse Response Functions
- (d) Autocorrelation Functions

- (e) Conditional expectations and variances
- (f) Companion Form
  - (i) Stationarity test
  - (ii) Formulae for means, variances, covariances

#### 6.4.4 Analogy Principle and Bias

- (a) How to apply the Analogy Principle
- (b) Formula for bias
- (c) Calculating bias
- (d) Bias-variance trade-off

Recall that here we are living in "expectations world."

#### 6.4.5 Asymptotics

- (a) Convergence in distribution
- (b) Convergence in probability
- (c) Weak Law of Large Numbers
- (d) Central Limit Theorem/Asymptotic Normality
- (e) Continuous Mapping Theorem
- (f) Slutsky's Theorem
- (g) The Delta Method

Now we are living in "asymptopia."<sup>3</sup>

---

<sup>3</sup>Check out this book on estimation.



# Chapter 7

## Asymptotics

### 7.1 Previous Problem: Gaussian Autoregressive Process

Consider the following autoregression model:

$$y_t = \mu + \phi(y_{t-1} - \mu) + \varepsilon_t \quad \varepsilon_t \stackrel{i.i.d}{\sim} N(0, \sigma^2)$$

#### 7.1.1 Part a

Calculate the expected value of the sample mean of  $y_t$ .

Using brute force:

$$\begin{aligned} \mathbb{E}[\bar{y}_T] &= \mathbb{E} \left[ \frac{1}{T} \sum_{t=1}^T y_t \right] \\ &= \frac{1}{T} \sum_{t=1}^T \mathbb{E}[y_t] \\ &= \frac{1}{T} \sum_{t=1}^T \mu \\ &= \mu \end{aligned}$$

#### 7.1.2 Part b

Write the AR(1) in its MA( $\infty$ ) representation. Provide an explicit description of the MA( $\infty$ ) coefficients  $\psi_j$ .

Using recursive substitution:

$$\begin{aligned}
y_t &= \mu + \phi(\mu + \phi(y_{t-2} - \mu) + \varepsilon_{t-1} - \mu) + \varepsilon_t \\
&= \mu + \phi^2(y_{t-2} - \mu) + \phi\varepsilon_{t-1} + \varepsilon_t \\
&= \mu + \phi^2(\mu + \phi(y_{t-3} - \mu) + \varepsilon_{t-2} - \mu) + \phi\varepsilon_{t-1} + \varepsilon_t \\
&= \mu + \phi^3(y_{t-3} - \mu) + \phi^2\varepsilon_{t-2} + \phi\varepsilon_{t-1} + \varepsilon_t
\end{aligned}$$

We can see the pattern and note that this process will become:

$$y_t = \mu + \sum_{j=1}^{\infty} \phi^j \varepsilon_{t-j}$$

Therefore the MA weights are  $\psi_j = \phi^j$ .

### 7.1.3 Part c

Consider the following sum:

$$S = \sum_{j=0}^{\infty} \psi_j$$

Provide an explicit expression for  $S$ .

This is fairly straightforward if we remember our geometric series:

$$\begin{aligned}
S &= \sum_{j=0}^{\infty} \psi_j \\
&= \sum_{j=0}^{\infty} \phi^j \\
&= \frac{1}{1 - \phi}
\end{aligned}$$

### 7.1.4 Part d

Consider the variance of the sample mean:

$$\text{Var}(\bar{y}_T) = \text{Var}\left(\frac{1}{T} \sum_{t=1}^T y_t\right)$$

Provide an explicit expression for the variance. How does this compare to when  $y_t$  is *i.i.d.*?

This is painful. But it is good to see at least once.

$$\begin{aligned}
\text{Var}(\bar{y}_T) &= \mathbb{E}[(\bar{y}_T - \mu)^2] \\
&= \mathbb{E}\left[\left(\frac{1}{T}\sum_{t=1}^T y_t - \mu\right)^2\right] \\
&= \mathbb{E}\left[\left(\frac{1}{T}\sum_{t=1}^T (y_t - \mu)\right)^2\right] \\
&= \frac{1}{T^2}\mathbb{E}[\{(y_1 - \mu) + (y_2 - \mu) + \dots + (y_T - \mu)\}\{(y_1 - \mu) + (y_2 - \mu) + \dots + (y_T - \mu)\}] \\
&= \frac{1}{T^2}\mathbb{E}[(y_1 - \mu)\{(y_1 - \mu) + (y_2 - \mu) + \dots + (y_T - \mu)\} \\
&\quad + (y_2 - \mu)\{(y_1 - \mu) + (y_2 - \mu) + \dots + (y_T - \mu)\} + \\
&\quad \dots + (y_T - \mu)\{(y_1 - \mu) + (y_2 - \mu) + \dots + (y_T - \mu)\}]
\end{aligned}$$

Notice that these are autocovariances. So:

$$\begin{aligned}
\text{Var}(\bar{y}_T) &= \frac{1}{T^2}[\gamma_0 + \gamma_1 + \gamma_2 + \dots + \gamma_{T-1} \\
&\quad + \gamma_1 + \gamma_0 + \gamma_1 + \dots + \gamma_{T-2} \\
&\quad + \dots + \gamma_{T-1} + \gamma_{T-2} + \gamma_{T-3} + \dots + \gamma_0] \\
&= \frac{1}{T^2}[T\gamma_0 + 2(T-1)\gamma_1 + 2(T-2)\gamma_2 + \dots + 2\gamma_{T-1}]
\end{aligned}$$

Now recall that for the AR(1) model,  $\gamma_j = \phi^j \gamma_0$  and  $\gamma_0 = \frac{\sigma^2}{1-\phi^2}$ . Plugging these in:

$$\begin{aligned}
\text{Var}(\bar{y}_T) &= \frac{1}{T^2}[T\gamma_0 + 2(T-1)\phi\gamma_0 + 2(T-2)\phi^2\gamma_0 + \dots + 2\phi^{T-1}\gamma_0] \\
&= \frac{1}{T^2}\gamma_0[T + 2(T-1)\phi + 2(T-2)\phi^2 + \dots + 2\phi^{T-1}] \\
&= \frac{\sigma^2}{T^2(1-\phi^2)}[T + 2(T-1)\phi + 2(T-2)\phi^2 + \dots + 2\phi^{T-1}]
\end{aligned}$$

This expression is the variance for  $\bar{y}_T$ . Note that if  $\phi = 0$ , then the process is *i.i.d.* and we get:

$$\text{Var}(\bar{y}_T) = \frac{\sigma^2}{T}$$

This expression is less than the the non-*i.i.d.* case.

### 7.1.5 Part e

Under the parametric assumption that the data is normally distributed, what is the exact finite sampling distribution of the sample mean  $\bar{y}_T$ ?

Sums of normal random variables are normal, so the sampling distribution is:

$$\bar{y}_T \sim N\left(\mu, \text{Var}\left(\sqrt{T} \cdot \bar{y}_T\right)\right)$$

## 7.2 Practice Problem 1: Hansen 8.1

Let  $X$  be distributed Bernoulli with  $P(X = 1) = p$  and  $P(X = 0) = 1 - p$ .

- (a) Show that  $p = \mathbb{E}[X]$
- (b) Write down the moment estimator  $\hat{p}$
- (c) Find  $\text{Var}(\hat{p})$
- (d) Find the asymptotic distribution of  $\sqrt{n}(\hat{p} - p)$  as  $n \rightarrow \infty$

### 7.2.1 Part a

We use the old definition of  $\mathbb{E}[X]$  from way back at the beginning of the semester:

$$\begin{aligned}\mathbb{E}[X] &= p(1) + (1 - p)(0) \\ &= p\end{aligned}$$

### 7.2.2 Part b

Let's use the analogy principle:

$$\begin{aligned}p &= \mathbb{E}[X] \\ \hat{p} &= \frac{1}{n} \sum_{i=1}^n x_i\end{aligned}$$

### 7.2.3 Part c

No tricks here either. We'll use brute force:



$$\begin{aligned}
 \text{Var}(\hat{p}) &= \text{Var}\left(\frac{1}{n} \sum_{i=1}^n x_i\right) && \text{Assuming independence:} \\
 &= \frac{1}{n^2} \sum_{i=1}^n \text{Var}(x_i) \\
 &= \frac{1}{n^2} \sum_{i=1}^n p(1-p) \\
 &= \frac{1}{n} p(1-p)
 \end{aligned}$$

### 7.2.4 Part d

Using the Central Limit Theorem:

$$\sqrt{n}(\hat{p} - p) \xrightarrow{d} N(0, V)$$

Where we know the mean asymptotically is zero because  $\hat{p}$  converges to  $p$  by the WLLN. There are two ways to find the asymptotic variance here. First is the brute force method:

$$\begin{aligned}
 V &= \mathbb{E}[(x - \mathbb{E}[x])^2] \\
 &= \mathbb{E}[x^2 - 2\mathbb{E}[x]x + \mathbb{E}[x]^2] \\
 &= \mathbb{E}[x^2] - 2\mathbb{E}[x]^2 + \mathbb{E}[x]^2 \\
 &= \mathbb{E}[x^2] - \mathbb{E}[x]^2 \\
 &= p - p^2 \\
 &= p(1-p)
 \end{aligned}$$

or second we can use:

$$\begin{aligned}
 V &= \text{Var}(\sqrt{n}\hat{p}) \\
 &= n\text{Var}(\hat{p}) \\
 &= n \cdot \frac{1}{n} p(1-p) \\
 &= p(1-p)
 \end{aligned}$$

Either method should always give the same answer.

### 7.3 Practice Problem 2: Hansen 8.3

Find the moment estimator of  $\mu'_3 = \mathbb{E}[x^3]$  and show that  $\sqrt{n}(\hat{\mu}'_3 - \mu'_3) \xrightarrow{d} N(0, V)$ . In addition, write  $V$  as a function of the moments of  $X$ .

#### 7.3.1 Solution

We first use the analogy principle:

$$\begin{aligned}\mu'_3 &= \mathbb{E}[x^3] \\ \hat{\mu}'_3 &= \frac{1}{n} \sum_{i=1}^n x_i^3\end{aligned}$$

Next, we show that  $\hat{\mu}'_3$  is consistent. This is straightforward, as the WLLN takes care of it:

$$\hat{\mu}'_3 = \frac{1}{n} \sum_{i=1}^n x_i^3 \xrightarrow{P} \mathbb{E}[x^3] = \mu'_3$$

So now we know that  $\sqrt{n}(\hat{\mu}'_3 - \mu'_3) \xrightarrow{d} N(0, V)$ . Next, we find  $V$ :

$$\begin{aligned}V &= \mathbb{E} \left[ (x^3 - \mathbb{E}[x^3])^2 \right] \\ &= \mathbb{E} \left[ x^6 - 2x^3 \mathbb{E}[x^3] + \mathbb{E}[x^3]^2 \right] \\ &= \mathbb{E}[x^6] - 2\mathbb{E}[x^3]^2 + \mathbb{E}[x^3]^2 \\ &= \mathbb{E}[x^6] - \mathbb{E}[x^3]^2\end{aligned}$$

We could have used the same trick as we used in the first practice problem if we wanted to calculate the sample variance for the estimator.

### 7.4 Practice Problem 3: Hansen 8.8

Assume that:

$$\sqrt{n} \begin{pmatrix} \hat{\theta}_1 - \theta_1 \\ \hat{\theta}_2 - \theta_2 \end{pmatrix} \xrightarrow{d} N(0, \Sigma)$$

Use the Delta Method to find the asymptotic distribution of the following statistics:

- (a)  $\hat{\theta}_1 \hat{\theta}_2$
- (b)  $e^{\hat{\theta}_1 + \hat{\theta}_2}$

(c) If  $\theta_2 \neq 0$ ,  $\frac{\hat{\theta}_1}{\hat{\theta}_2^2}$

(d)  $\hat{\theta}_1^3 + \hat{\theta}_1 \hat{\theta}_2^2$

### 7.4.1 Part a

Using Slutsky's Theorem, we know that  $\hat{\theta}_1 \hat{\theta}_2 \xrightarrow{P} \theta_1 \theta_2$  since  $\hat{\theta}_1 \xrightarrow{P} \theta_1$  and  $\hat{\theta}_2 \xrightarrow{P} \theta_2$  from the set-up of the problem. So we only need to find the asymptotic variance. Using the Delta Method, we start with  $H$ :

$$\begin{aligned} H &= \frac{\partial}{\partial \boldsymbol{\theta}} \theta_1 \theta_2 \\ &= \begin{bmatrix} \theta_2 \\ \theta_1 \end{bmatrix} \end{aligned}$$

We are given that  $V = \Sigma$ , so the asymptotic distribution is:

$$\sqrt{n}(\hat{\theta}_1 \hat{\theta}_2 - \theta_1 \theta_2) \xrightarrow{d} N \left( 0, \begin{bmatrix} \theta_2 \\ \theta_1 \end{bmatrix}^T \Sigma \begin{bmatrix} \theta_2 \\ \theta_1 \end{bmatrix} \right)$$

### 7.4.2 Part b

We know that  $e^{\hat{\theta}_1 + \hat{\theta}_2} \xrightarrow{P} e^{\theta_1 + \theta_2}$  by the Continuous Mapping Theorem since  $\hat{\theta}_1 + \hat{\theta}_2 \xrightarrow{P} \theta_1 + \theta_2$  by Slutsky's Theorem and the exponential transformation is continuous. We now find  $H$ :

$$\begin{aligned} H &= \frac{\partial}{\partial \boldsymbol{\theta}} e^{\theta_1 + \theta_2} \\ &= \begin{bmatrix} e^{\theta_1 + \theta_2} \\ e^{\theta_1 + \theta_2} \end{bmatrix} \end{aligned}$$

Putting this altogether:

$$\sqrt{n} \left( e^{\hat{\theta}_1 + \hat{\theta}_2} - e^{\theta_1 + \theta_2} \right) \xrightarrow{d} N \left( 0, \begin{bmatrix} e^{\theta_1 + \theta_2} \\ e^{\theta_1 + \theta_2} \end{bmatrix}^T \Sigma \begin{bmatrix} e^{\theta_1 + \theta_2} \\ e^{\theta_1 + \theta_2} \end{bmatrix} \right)$$

### 7.4.3 Part c

We know that  $\frac{\hat{\theta}_1}{\hat{\theta}_2^2} \xrightarrow{P} \frac{\theta_1}{\theta_2^2}$  by the CMT, as division is continuous as long as the denominator is not zero (which we are given), since by the CMT  $\hat{\theta}_2^2 \xrightarrow{P} \theta_2^2$ . We now find  $H$ :

$$\begin{aligned}
 H &= \frac{\partial}{\partial \boldsymbol{\theta}} \frac{\theta_1}{\theta_2^2} \\
 &= \begin{bmatrix} \frac{1}{\theta_2^2} \\ -\frac{2\theta_1}{\theta_2^3} \end{bmatrix}
 \end{aligned}$$

Altogether, we now have:

$$\sqrt{n} \left( \frac{\hat{\theta}_1}{\hat{\theta}_2^2} - \frac{\theta_1}{\theta_2^2} \right) \xrightarrow{d} N \left( 0, \begin{bmatrix} \frac{1}{\theta_2^2} \\ -\frac{2\theta_1}{\theta_2^3} \end{bmatrix}^T \Sigma \begin{bmatrix} \frac{1}{\theta_2^2} \\ -\frac{2\theta_1}{\theta_2^3} \end{bmatrix} \right)$$

#### 7.4.4 Part d

We know that  $\hat{\theta}_1^3 + \hat{\theta}_1 \hat{\theta}_2^2 \xrightarrow{P} \theta_1^3 + \theta_1 \theta_2^2$  by the CMT since  $\hat{\theta}_1^3 \xrightarrow{P} \theta_1^3$  by the CMT,  $\hat{\theta}_2^2 \xrightarrow{P} \theta_2^2$  by the CMT, and  $\hat{\theta}_1 \hat{\theta}_2^2 \xrightarrow{P} \theta_1 \theta_2^2$  by Slutsky's theorem. We now find  $H$ :

$$\begin{aligned}
 H &= \frac{\partial}{\partial \boldsymbol{\theta}} (\theta_1^3 + \theta_1 \theta_2^2) \\
 &= \begin{bmatrix} 3\theta_1^2 + \theta_2^2 \\ 2\theta_1 \theta_2 \end{bmatrix}
 \end{aligned}$$

Now pulling this all together:

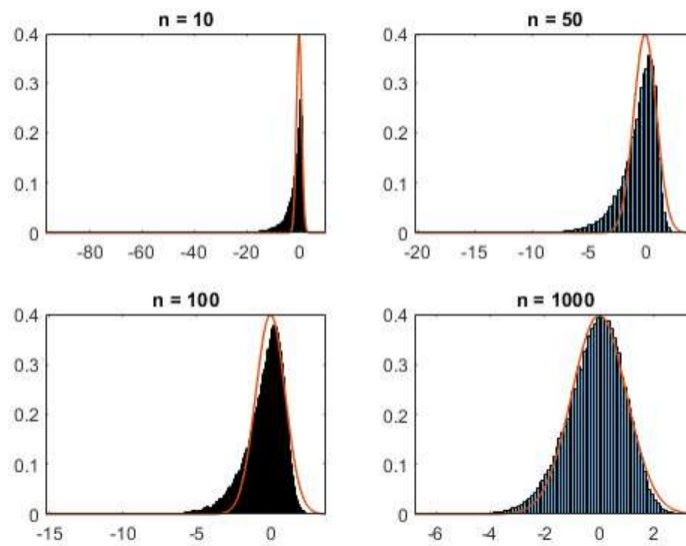
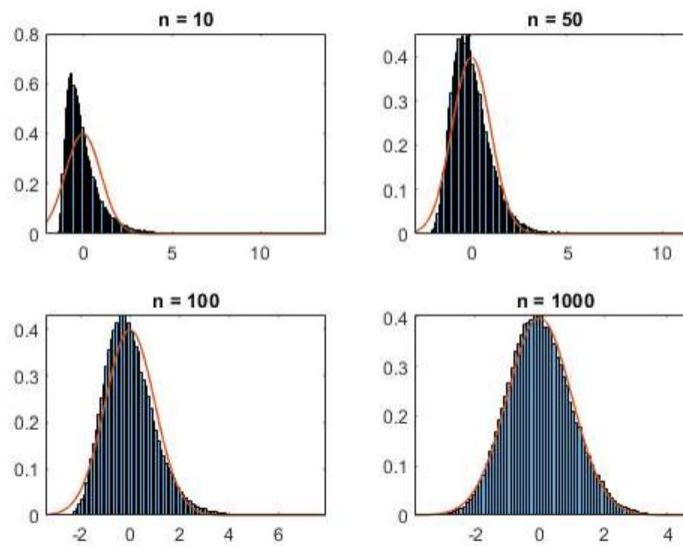
$$\sqrt{n} \left( \hat{\theta}_1^3 + \hat{\theta}_1 \hat{\theta}_2^2 - (\theta_1^3 + \theta_1 \theta_2^2) \right) \xrightarrow{d} N \left( 0, \begin{bmatrix} 3\theta_1^2 + \theta_2^2 \\ 2\theta_1 \theta_2 \end{bmatrix}^T \Sigma \begin{bmatrix} 3\theta_1^2 + \theta_2^2 \\ 2\theta_1 \theta_2 \end{bmatrix} \right)$$

## 7.5 Matlab Checkpoints

The Matlab comes at the end of a long theoretical exercise. Make sure you stay close to the theorems and you'll be fine. To help you out, here are a few checks to make sure your code's output is correct.

First, if you use parameter values of  $\alpha = 2$  and  $\delta = 0.5$ , your true asymptotic variance *for your estimator* should be 1.25.

Second, here are some pictures of how your histograms should eventually look:

Asymptotic Normality of  $\hat{\sigma}^2$ , Estimated VarianceAsymptotic Normality of  $\hat{\sigma}^2$ , True Variance

Each histogram has 50,000 different variance calculations. Hopefully these help you stay on the right track.



## Chapter 8

# Maximum Likelihood Estimation

### 8.1 Maximum Likelihood Theory

#### 8.1.1 Basics of MLE

The Maximum Likelihood Estimator seeks to estimate a parameter that maximizes the likelihood function. The likelihood function describes the probability of observing the data that we've collected with parameters as the arguments. Of course, the underlying functions and therefore parameters in use are chosen by the modeller. The likelihood function is given as:

$$L_n(\boldsymbol{\theta}|\mathbf{Y}) = \prod_{i=1}^n f(\mathbf{y}_i|\boldsymbol{\theta})$$

where  $f(\mathbf{y}_i|\boldsymbol{\theta})$  is the underlying DGP of the data.

To make analysis easier, we often take the natural log of the likelihood function. Note that because natural log is a monotonically increasing function, the argmax of the log transformation is the same as the argmax of the original likelihood function.

The log-likelihood function is:

$$\ell_n(\boldsymbol{\theta}) = \sum_{i=1}^n \ln(f(\mathbf{y}_i|\boldsymbol{\theta}))$$

Remember that these functions are functions of  $\boldsymbol{\theta}$  and we keep  $\mathbf{y}$  fixed.

To estimate  $\boldsymbol{\theta}$ , we find the score vector. The score vector is defined as:

$$\mathbf{s}_n(\boldsymbol{\theta}) = \frac{\partial \ell_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$$

Note that if  $\boldsymbol{\theta}$  is  $k \times 1$ , then  $\mathbf{s}_n(\boldsymbol{\theta})$  is also  $k \times 1$ .

If the log-likelihood function is differentiable over  $\boldsymbol{\theta}$ , then we can set the score vector equal to zero and solve for  $\hat{\boldsymbol{\theta}}$ . Recall from your microeconomics class that this FOC is necessary and not sufficient for determining whether we are at a maximum. As such, we should check SOSCs ( $\frac{\partial^2 \ell_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} < 0$ ) just to be sure.

### 8.1.2 Fisher Information Matrix

The Fisher information matrix conveys how much information the data  $\mathbf{Y}$  carries about the unknown parameter  $\boldsymbol{\theta}$ . It is given by:

$$\begin{aligned} \mathcal{I}_{\boldsymbol{\theta},n} &= \text{Var} \left( \frac{\partial \ell_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) \\ &= \mathbb{E} \left[ \frac{\partial \ell_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \ell_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \right] \end{aligned}$$

Note that if our model is regular (see slide 23 of Topic 7) and correctly specified (see slide 6 of Topic 7), then the Fisher information matrix equals the expected Hessian:

$$\begin{aligned} \mathcal{I}_{\boldsymbol{\theta},n} &= \mathcal{H}_{\boldsymbol{\theta},n} \\ \mathbb{E} \left[ \frac{\partial \ell_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \ell_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \right] &= \mathbb{E} \left[ -\frac{\partial^2 \ell_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right] \end{aligned}$$

It is useful to remember that the asymptotic distribution of our MLE estimators is:

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{d} N \left( 0, \mathcal{I}_{\boldsymbol{\theta},1}^{-1} \right)$$

as long as assumptions 1-11 on slide 54 of Topic 7 hold.

If the model is mis-specified, then the asymptotic distribution is:

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{d} N \left( 0, \mathcal{H}_{\boldsymbol{\theta},1}^{-1} \mathcal{I}_{\boldsymbol{\theta},1} \mathcal{H}_{\boldsymbol{\theta},1}^{-1} \right)$$

### 8.1.3 Cramér-Rao Lower Bound

If a model is regular, parametric, and correctly specified with an interior solution for  $\boldsymbol{\theta}$  that is unbiased, then the lowest possible variance is the Cramér-Rao Lower Bound. That is:

$$\text{Var}(\hat{\boldsymbol{\theta}}_n) \geq (n\mathcal{I}_{\boldsymbol{\theta},1})^{-1}$$

Note that if our data is *i.i.d.*, then  $n\mathcal{I}_{\boldsymbol{\theta},1} = \mathcal{I}_{\boldsymbol{\theta},n}$ .



### 8.1.4 Estimating the Asymptotic Variance of the Estimator

I'm not going to type all of these out, but you can find the four different ways to estimate the asymptotic variance at the end of the Topic 7 slides. Pay close attention to the definitions as the differences between the estimates can be difficult to understand.

## 8.2 Practice Problem 1: Hansen 10.4

Let  $X$  be distributed Pareto with density  $f(x) = \frac{1}{\pi(1+(x-\theta)^2)}$  for  $x \in \mathbb{R}$ .

- Find the log-likelihood function of  $\ell_n(\theta)$ .
- Find the first-order condition for the MLE  $\hat{\theta}$  for  $\theta$ . You will not be able to solve for  $\hat{\theta}$ .

### 8.2.1 Part a

We begin by finding the likelihood function. We will then take the natural log of that function. Using the definition of the likelihood function:

$$\begin{aligned}\mathcal{L}(\theta|x) &= \prod_{i=1}^n \frac{1}{\pi(1+(x_i-\theta)^2)} \\ \ell_n(\theta) &= \sum_{i=1}^n -\ln(\pi(1+(x_i-\theta)^2)) \\ &= \sum_{i=1}^n -\ln(\pi) - \ln(1+(x_i-\theta)^2) \\ &= -n\ln(\pi) - \sum_{i=1}^n \ln(1+(x_i-\theta)^2)\end{aligned}$$

### 8.2.2 Part b

Now that we have the log-likelihood function, we can find the score vector and set that equal to zero:

$$\begin{aligned}\frac{\partial \ell_n(\theta)}{\partial \theta} &= \frac{\partial}{\partial \theta} \left[ -n\ln(\pi) - \sum_{i=1}^n \ln(1+(x_i-\theta)^2) \right] \\ &= \sum_{i=1}^n \frac{2(x_i-\theta)}{1+(x_i-\theta)^2} \\ &= 0\end{aligned}$$

Usually we'd like to solve for  $\hat{\theta}$ , but note that here we cannot. We need a numerical solver to find the optimal estimator. So our work, analytically at least, is done.

### 8.3 Practice Problem 2: Hansen 10.7

Take the Pareto model  $f(x) = \alpha x^{-1-\alpha}$ ,  $x \geq 1$ . Calculate the information for  $\alpha$  using the second derivative.

#### 8.3.1 Solution

We start with the likelihood function. We will then take the natural log to find the log-likelihood function.

$$\begin{aligned}\mathcal{L}(\alpha|x) &= \prod_{i=1}^n (\alpha x_i^{-1-\alpha}) \\ \ell_n(\alpha) &= \sum_{i=1}^n \ln(\alpha x_i^{-1-\alpha}) \\ &= \sum_{i=1}^n \ln(\alpha) - (1+\alpha)\ln(x_i) \\ &= n\ln(\alpha) - \sum_{i=1}^n (1+\alpha)\ln(x_i)\end{aligned}$$

Now that we have the log-likelihood we can find the score vector:

$$\begin{aligned}\frac{\partial \ell_n(\alpha)}{\partial \alpha} &= \frac{\partial}{\partial \alpha} \left[ n\ln(\alpha) - \sum_{i=1}^n (1+\alpha)\ln(x_i) \right] \\ &= \frac{n}{\alpha} - \sum_{i=1}^n \ln(x_i)\end{aligned}$$

The problem tells us to use the second derivative, so taking the derivative with respect to  $\alpha$  again:

$$\begin{aligned}\frac{\partial^2 \ell_n(\alpha)}{\partial \alpha^2} &= \frac{\partial}{\partial \alpha} \left[ \frac{n}{\alpha} - \sum_{i=1}^n \ln(x_i) \right] \\ &= -\frac{n}{\alpha^2}\end{aligned}$$

Now that we have the second derivative, we look at the formula for the expected Hessian:

$$\begin{aligned}\mathcal{H}_{\theta,n} &= \mathbb{E} \left[ -\frac{\partial^2 \ell_n(\alpha)}{\partial \alpha^2} \right] \\ &= \mathbb{E} \left[ \frac{n}{\alpha^2} \right] \\ &= \frac{n}{\alpha^2}\end{aligned}$$

Assuming the information matrix equality holds, then:

$$\mathcal{I}_{\theta,n} = \frac{n}{\alpha^2}$$

## 8.4 Practice Problem 3: Hansen 10.8

Find the Cramér-Rao lower bound for  $p$  in the Bernoulli model. In Section 10.3, we derived that the MLE for  $p$  is  $\hat{p} = \bar{X}_n$ . Compute  $Var(\hat{p})$ . Compare  $Var(\hat{p})$  with the Cramér-Rao lower bound.

### 8.4.1 Finding the Cramér-Rao lower bound

We first note that the Cramér-Rao lower bound is given by  $(n\mathcal{I}_{\theta,1})^{-1}$ . So we know we need to first find the likelihood function. The pmf of a Bernoulli random variable is  $\pi(x) = p^x(1-p)^{1-x}$ . So:

$$\begin{aligned}\mathcal{L}(p|x) &= \prod_{i=1}^n p^{x_i}(1-p)^{1-x_i} \\ \ell_n(p) &= \sum_{i=1}^n \ln(p^{x_i}(1-p)^{1-x_i}) \\ &= \sum_{i=1}^n x_i \ln(p) + (1-x_i) \ln(1-p)\end{aligned}$$

Next, we know we need the score vector, so:

$$\begin{aligned}\frac{\partial \ell_n(p)}{\partial p} &= \frac{\partial}{\partial p} \left[ \sum_{i=1}^n x_i \ln(p) + (1-x_i) \ln(1-p) \right] \\ &= \frac{\sum_{i=1}^n x_i}{p} - \frac{\sum_{i=1}^n (1-x_i)}{1-p}\end{aligned}$$

If we assume that the information matrix equality holds (which you'll show in Hansen 10.6), then we can use the second derivative:

$$\begin{aligned}\frac{\partial^2 \ell_n(p)}{\partial p^2} &= \frac{\partial}{\partial p} \left[ \frac{\sum_{i=1}^n x_i}{p} - \frac{\sum_{i=1}^n (1-x_i)}{1-p} \right] \\ &= -\frac{\sum_{i=1}^n x_i}{p^2} - \frac{\sum_{i=1}^n (1-x_i)}{(1-p)^2}\end{aligned}$$

Now take the negative expectation:

$$\begin{aligned}
\mathbb{E} \left[ -\frac{\partial^2 \ell_n(p)}{\partial p^2} \right] &= \mathbb{E} \left[ \frac{\sum_{i=1}^n x_i}{p^2} + \frac{\sum_{i=1}^n (1-x_i)}{(1-p)^2} \right] \\
&= \frac{1}{p^2} \sum_{i=1}^n \mathbb{E}[x_i] + \frac{n}{(1-p)^2} - \frac{\sum_{i=1}^n \mathbb{E}[x_i]}{(1-p)^2} \\
&= \frac{np}{p^2} + \frac{n}{(1-p)^2} - \frac{np}{(1-p)^2} \\
&= \frac{n}{p} + \frac{n(1-p)}{(1-p)^2} \\
&= \frac{n - np + np}{p(1-p)} \\
&= \frac{n}{p(1-p)}
\end{aligned}$$

Since the information matrix equality is assumed to hold (again, you'll show this in Hansen 10.6):

$$\mathcal{I}_{\theta,n} = \frac{n}{p(1-p)}$$

and because our data is *i.i.d.*:

$$n\mathcal{I}_{\theta,1} = \frac{n}{p(1-p)}$$

We have one last step. We must invert this last expression:

$$(n\mathcal{I}_{\theta,1})^{-1} = \frac{p(1-p)}{n}$$

This is the Cramér-Rao lower bound.

### 8.4.2 Variance of the Estimator

We still need to find  $\text{Var}(\hat{p})$ . So:

$$\begin{aligned}
 \text{Var}(\hat{p}) &= \text{Var}\left(\frac{1}{n} \sum_{i=1}^n x_i\right) \\
 &= \frac{1}{n^2} \text{Var}\left(\sum_{i=1}^n x_i\right) \\
 &= \frac{1}{n^2} \sum_{i=1}^n \text{Var}(x_i) \\
 &= \frac{1}{n^2} \sum_{i=1}^n p(1-p) \\
 &= \frac{n}{n^2} p(1-p) \\
 &= \frac{p(1-p)}{n}
 \end{aligned}$$

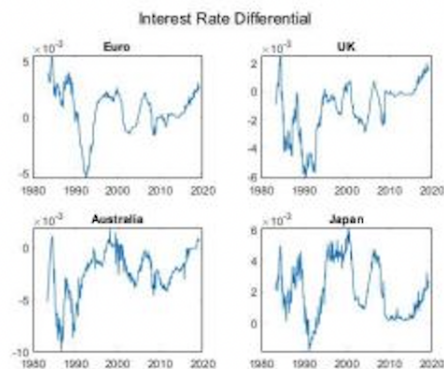
Note that this is the Cramér-Rao lower bound.

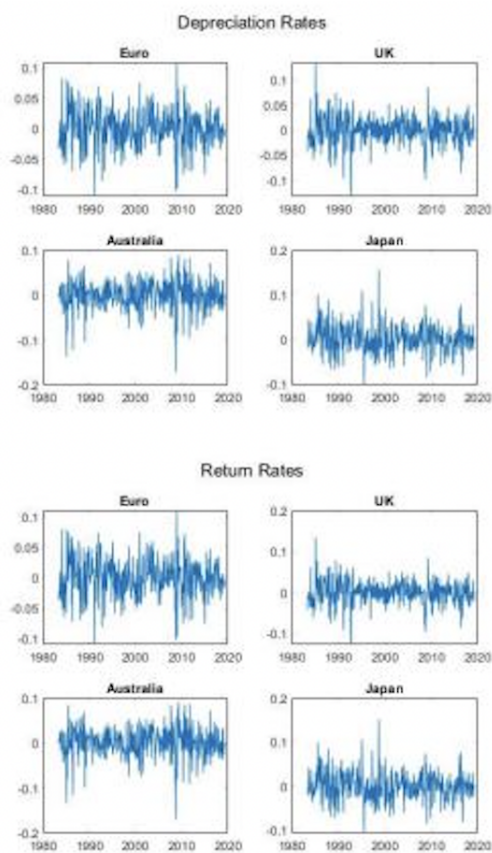
## 8.5 Matlab Help

The Matlab this week is tricky again. The analytical part is fairly straightforward - just be sure to check your algebra carefully! As a hint: define your data in terms of sample means to make the math more tractable.

On part v, you are asked to derive estimates for the asymptotic covariance matrices. Another hint: you should arrive at the same matrix for both  $\hat{V}_0$  and  $\hat{V}_1$ . If you don't, go back and rework the problem.

For the coding section, here are pictures of what your time series plots should look like:



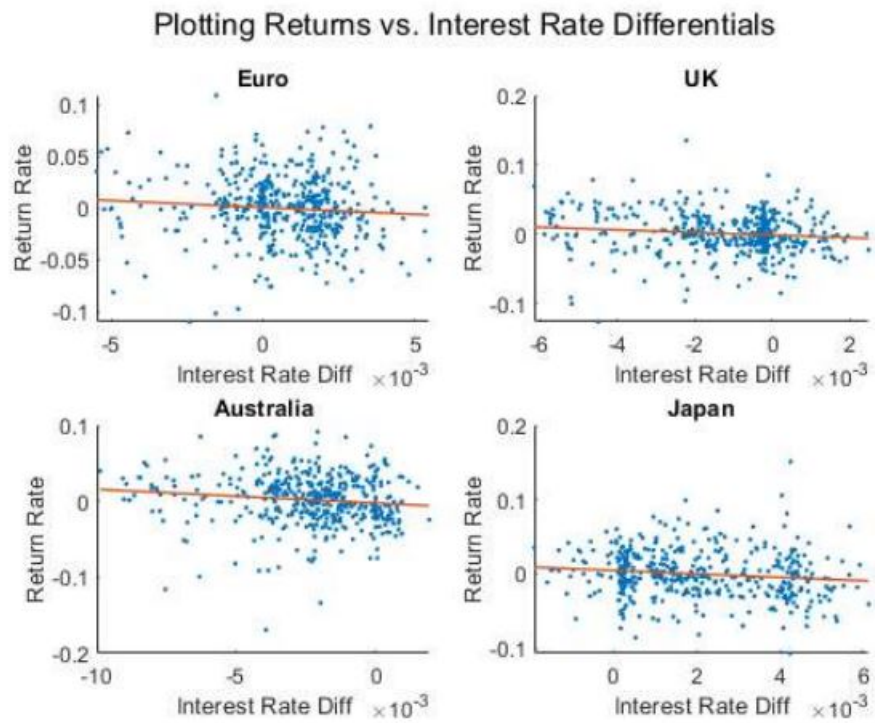


As another checkpoint, your robust variance-covariance matrix for the Euro-zone should look like this:

0.0012	-0.2628	-7.8928e-15
-0.2628	327.1666	5.9747e-12
-7.8928e-15	5.9747e-12	9.1492e-07

with your variances for  $\hat{\alpha}$ ,  $\hat{\beta}$ , and  $\hat{\sigma}^2$  respectively down the diagonal.

Lastly, the very last section asks you to plot a regression line over the raw data for interest rate differentials versus return rates. Your end result should look like:



Hopefully these checkpoints and hints are helpful. Good luck on the problem set.





# Chapter 9

## Method of Moments

### 9.1 Previous Problem 1: Matlab Theory Part

In the previous problem set, Drew gave us a linear regression model:

$$y_i = \alpha + \beta x_i + \varepsilon_i \quad \varepsilon_i \sim N(0, \sigma^2)$$

where  $y_i \sim N(\alpha + \beta x_i, \sigma^2)$ . Since  $y_i$  is distributed normally, we know its pdf:

$$f(y_i|x_i) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y_i - \alpha - \beta x_i)^2}$$

Conditioning on the  $x_i$ , we want to find the Fisher Information matrix.

#### 9.1.1 Conditional Log-Likelihood

Start first with the likelihood, then take the log:

$$\begin{aligned} L_n(\boldsymbol{\theta}) &= \prod_{i=1}^n f(y_i|x_i) \\ \ell_n(\boldsymbol{\theta}) &= \sum_{i=1}^n \ln \left( \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y_i - \alpha - \beta x_i)^2} \right) \\ &= -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2 \end{aligned}$$

#### 9.1.2 Score Vector

Next we take the derivative of the log-likelihood with respect to each of our three parameters:

$$\begin{aligned}\frac{\partial \ell_n(\boldsymbol{\theta})}{\partial \alpha} &= \frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \alpha - \beta x_i) \\ \frac{\partial \ell_n(\boldsymbol{\theta})}{\partial \beta} &= \frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \alpha - \beta x_i) x_i \\ \frac{\partial \ell_n(\boldsymbol{\theta})}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2\end{aligned}$$

Let's simplify one equation at a time, starting with  $\alpha$ :

$$\begin{aligned}\frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \alpha - \beta x_i) &= \frac{1}{\sigma^2} \left[ \sum_{i=1}^n y_i - n\alpha - \beta \sum_{i=1}^n x_i \right] \\ &= \frac{1}{\sigma^2} [n\bar{y} - n\alpha - \beta n\bar{x}]\end{aligned}$$

Recall that we set these first-order conditions equal to zero, so we can multiply by  $\sigma^2$  and add  $n\alpha$  to get:

$$\begin{aligned}n\alpha &= n\bar{y} - n\beta\bar{x} \\ \hat{\alpha} &= \bar{y} - \beta\bar{x}\end{aligned}$$

Before we can advance further, we must solve for  $\hat{\beta}$ :

$$\begin{aligned}0 &= \frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \alpha - \beta x_i) x_i \\ &= \frac{1}{\sigma^2} \sum_{i=1}^n x_i y_i - \alpha x_i - \beta x_i^2 \\ &= \frac{1}{\sigma^2} [n\bar{x}\bar{y} - n\alpha\bar{x} - n\beta\bar{x}\bar{x}] \\ \hat{\beta} &= \frac{\bar{x}\bar{y} - \alpha\bar{x}}{\bar{x}\bar{x}}\end{aligned}$$

Now we can plug  $\hat{\beta}$  into  $\hat{\alpha}$ :

$$\begin{aligned}\hat{\alpha} &= \bar{y} - \left( \frac{\overline{xy} - \alpha \bar{x}}{\overline{xx}} \right) \bar{x} \\ \hat{\alpha} \left( 1 - \frac{\bar{x}^2}{\overline{xx}} \right) &= \bar{y} - \frac{\overline{xy}}{\overline{xx}} \\ \hat{\alpha} (\overline{xx} - \bar{x}^2) &= \overline{xy} - \overline{xy} \bar{x} \\ \hat{\alpha} &= \frac{\overline{xy} - \overline{xy} \bar{x}}{\overline{xx} - \bar{x}^2}\end{aligned}$$

We can then plug  $\hat{\alpha}$  back into  $\hat{\beta}$ :

$$\begin{aligned}\hat{\beta} &= \frac{\overline{xy} - \frac{\overline{xy} - \overline{xy} \bar{x}}{\overline{xx} - \bar{x}^2} \bar{x}}{\overline{xx}} \\ &= \frac{\overline{xy}(\overline{xx} - \bar{x}^2) - \overline{xy} \bar{x} + \overline{xy} \bar{x}^2}{\overline{xx}(\overline{xx} - \bar{x}^2)} \\ &= \frac{\overline{xy} - \overline{xy} \bar{x}}{\overline{xx} - \bar{x}^2}\end{aligned}$$

Now we can solve for  $\sigma^2$ :

$$\begin{aligned}0 &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2 \\ \frac{n}{2\sigma^2} &= \frac{1}{2\sigma^4} \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2 \\ \hat{\sigma}^2 &= \frac{1}{n} \sum_{i=1}^n (y_i - \hat{\alpha} - \hat{\beta} x_i)^2\end{aligned}$$

### 9.1.3 Deriving the Observed Hessian

The observed Hessian  $\mathcal{H}_{\boldsymbol{\theta},n}^o$  is defined as:

$$\mathcal{H}_{\boldsymbol{\theta},n}^o = \left[ \frac{\partial^2 \ell_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right]$$

Let's go one at a time:

$$\begin{aligned}
\frac{\partial^2 \ell_n(\boldsymbol{\theta})}{\partial \alpha^2} &= -\frac{n}{\sigma^2} \\
\frac{\partial^2 \ell_n(\boldsymbol{\theta})}{\partial \alpha \partial \beta} &= -\frac{1}{\sigma^2} \sum_{i=1}^n x_i \\
\frac{\partial^2 \ell_n(\boldsymbol{\theta})}{\partial \alpha \partial \sigma^2} &= -\frac{1}{\sigma^4} \sum_{i=1}^n (y_i - \alpha - \beta x_i) \\
\frac{\partial^2 \ell_n(\boldsymbol{\theta})}{\partial \beta^2} &= -\frac{1}{\sigma^2} \sum_{i=1}^n x_i^2 \\
\frac{\partial^2 \ell_n(\boldsymbol{\theta})}{\partial \beta \partial \sigma^2} &= -\frac{1}{\sigma^4} \sum_{i=1}^n (y_i - \alpha - \beta x_i) x_i \\
\frac{\partial^2 \ell_n(\boldsymbol{\theta})}{(\partial \sigma^2)^2} &= \frac{n}{2\sigma^4} - \frac{1}{\sigma^6} \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2
\end{aligned}$$

#### 9.1.4 Information Matrix

If we assume the information matrix equality holds, then the expectation of the negative of the observed Hessian is the same as the information matrix. Recall that we treat  $x_i$  as fixed data so that its mean is its expectation. Let's first take the negative expectation of the second derivatives we calculated:

$$\begin{aligned}
\mathbb{E} \left[ -\frac{\partial^2 \ell_n(\boldsymbol{\theta})}{\partial \alpha^2} \right] &= \frac{n}{\sigma^2} \\
\mathbb{E} \left[ -\frac{\partial^2 \ell_n(\boldsymbol{\theta})}{\partial \alpha \partial \beta} \right] &= \frac{n}{\sigma^2} \bar{x} \\
\mathbb{E} \left[ -\frac{\partial^2 \ell_n(\boldsymbol{\theta})}{\partial \alpha \partial \sigma^2} \right] &= 0 \\
\mathbb{E} \left[ -\frac{\partial^2 \ell_n(\boldsymbol{\theta})}{\partial \beta^2} \right] &= \frac{n}{\sigma^2} \overline{xx} \\
\mathbb{E} \left[ -\frac{\partial^2 \ell_n(\boldsymbol{\theta})}{\partial \beta \partial \sigma^2} \right] &= 0 \\
\mathbb{E} \left[ -\frac{\partial^2 \ell_n(\boldsymbol{\theta})}{(\partial \sigma^2)^2} \right] &= \frac{n}{2\sigma^4}
\end{aligned}$$

Note that we use both the property that the sum of the residuals is zero and that the sum of squared residuals is equal to the variance.

So the information matrix is:

$$\mathcal{I}_{\boldsymbol{\theta}, n} = \begin{bmatrix} \frac{n}{\sigma^2} & \frac{n}{\sigma^2} \bar{x} & 0 \\ \frac{n}{\sigma^2} \bar{x} & \frac{n}{\sigma^2} \overline{xx} & 0 \\ 0 & 0 & \frac{n}{2\sigma^4} \end{bmatrix}$$

So that was a lot of work! Know how to calculate information matrices quickly for the comprehensive exam in June.

## 9.2 Method of Moments Theory

I will focus on the modern view of the Method of Moments, as Drew refers to it:

Let  $\boldsymbol{\beta}$  be a  $k \times 1$  vector of unknown parameters and  $\mathbf{m}(\cdot)$  be a  $k \times 1$  vector of functions that satisfy the population moment

$$\mathbb{E}[\mathbf{m}(\mathbf{y}, \boldsymbol{\beta})] = 0$$

The method of moments estimation  $\hat{\boldsymbol{\beta}}_n$  is the solution to this system of  $k$  unknowns in  $k$  equations (also known as the solution to the moment conditions):

$$\frac{1}{n} \sum_{i=1}^n \mathbf{m}(\mathbf{y}_i, \hat{\boldsymbol{\beta}}_n) = 0$$

This should look familiar! After all, we are basically using the analogy principle again.

A very simple example: estimating the population mean. Our moment condition is thus:

$$\begin{aligned} \mathbb{E}[y_i - \mu] &= 0 && \text{Pushing the expectation through:} \\ \mathbb{E}[y_i] - \mu &= 0 && \text{Adding the population moment across:} \\ \mathbb{E}[y_i] &= \mu \end{aligned}$$

Now we use the definition of the method of moments estimator:

$$\frac{1}{n} \sum_{i=1}^n y_i = \hat{\mu}$$

Which is exactly like the analogy principle. It may be tempting to jump straight to the analogy principle but really slow down and think through constructing the moment conditions.

Now suppose that we have a function of population moments:

$$\boldsymbol{\beta} = \mathbf{h}(\mathbb{E}[\mathbf{g}(\mathbf{y})])$$

We need to construct moment conditions for each population moment and for the function  $\boldsymbol{\beta}$ . So:

$$\mathbb{E}[\mathbf{m}(\mathbf{y}, \boldsymbol{\beta}, \boldsymbol{\theta})] = \begin{bmatrix} \mathbb{E}[\mathbf{g}(\mathbf{y})] - \boldsymbol{\theta} \\ \mathbf{h}(\boldsymbol{\theta}) - \boldsymbol{\beta} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Here, we would find our MoM estimators for each element of  $\theta$  and then plug them into  $\mathbf{h}$  to obtain our MoM plug-in estimator for  $\beta$ .

### 9.3 Practice Problem 1: Hansen 11.3

A Bernoulli random variable  $X$  is

$$\begin{aligned} P(X = 0) &= 1 - p \\ p(X = 1) &= p \end{aligned}$$

- (a) Propose a moment estimator  $\hat{p}$  for  $p$ .
- (b) Find the variance of the asymptotic distribution of  $\sqrt{n}(\hat{p} - p)$ .
- (c) Propose an estimator of the asymptotic variance of  $\hat{p}$ .

#### 9.3.1 Part a

We begin with the moment conditions:

$$\begin{aligned} \mathbb{E}[x_i - \mu] &= 0 \\ \mathbb{E}[x_i] &= \mu \end{aligned}$$

Note that for a Bernoulli random variable  $\mu = p$ . Use the method of moments estimator:

$$\begin{aligned} \hat{\mu} &= \frac{1}{n} \sum_{i=1}^n x_i \\ \hat{p} &= \frac{1}{n} \sum_{i=1}^n x_i \end{aligned}$$

#### 9.3.2 Part b

Using the WLLN, we know that  $\hat{p} \xrightarrow{P} p$ . So we just need to find the variance:

$$\begin{aligned} V &= \mathbb{E}[(x - p)(x - p)] \\ &= \mathbb{E}[(x - p)^2] \\ &= p(1 - p) \end{aligned}$$

#### 9.3.3 Part c

This part is really as easy as it seems. Put hats on top of the unknown parameters in our asymptotic variance:

$$\widehat{AVAR} = \hat{p}(1 - \hat{p})$$

## 9.4 Practice Problem 2: Hansen 11.4

Propose a moment estimator  $\hat{\lambda}$  for the parameter  $\lambda$  of a Poisson distribution.

### 9.4.1 Solution

Set-up our moment condition:

$$\begin{aligned}\mathbb{E}[x_i - \mu] &= 0 \\ \mathbb{E}[x_i] &= \mu\end{aligned}$$

Using the definition of the MoM estimator:

$$\frac{1}{n} \sum_{i=1}^n x_i = \hat{\mu}$$

Note that for the Poisson distribution  $\mu = \lambda$ , so:

$$\frac{1}{n} \sum_{i=1}^n x_i = \hat{\lambda}$$

## 9.5 Matlab Help

This is another week of brutal Matlab coding. This time, the homework asks you to code up a Newey-West Variance estimator and run Monte Carlo simulations with it. The Newey-West estimator is defined as follows:

$$\hat{\mathbf{V}}_{NW} = \hat{\mathbf{\Gamma}}_0 + \sum_{j=1}^q k(j) \cdot (\hat{\mathbf{\Gamma}}_j + \hat{\mathbf{\Gamma}}_j')$$

where  $q$  is defined as the number of lags included and  $k(j) = 1 - \frac{j}{q+1}$ . Of course, we do not know the true autocovariance matrices, so we estimate them using:

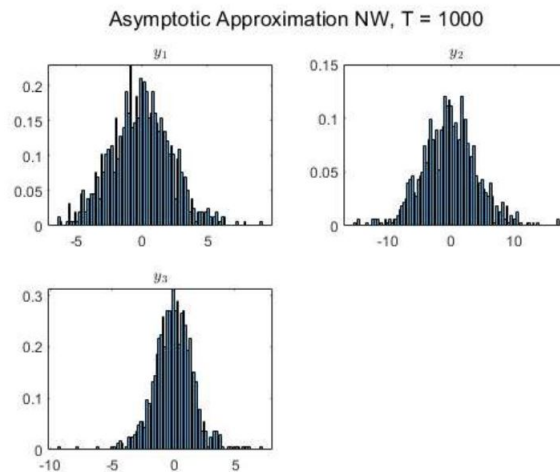
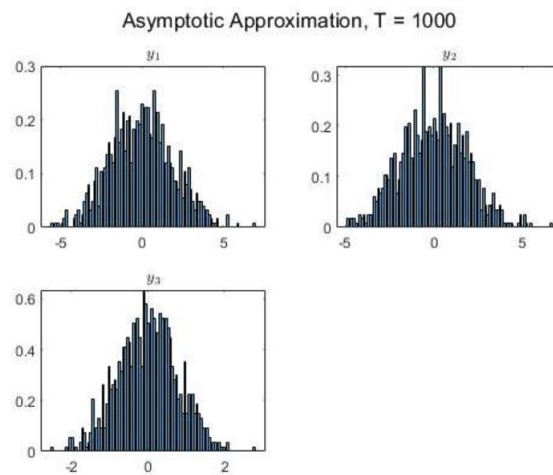
$$\hat{\mathbf{\Gamma}}_j = \frac{1}{T-1} \sum_{t=j+1}^T (\mathbf{y}_t - \bar{\mathbf{y}})(\mathbf{y}_{t-j} - \bar{\mathbf{y}})'$$

where  $T$  is the number of time-series observations. Pay close attention to the indices on the summations and the subscripts on the variables.

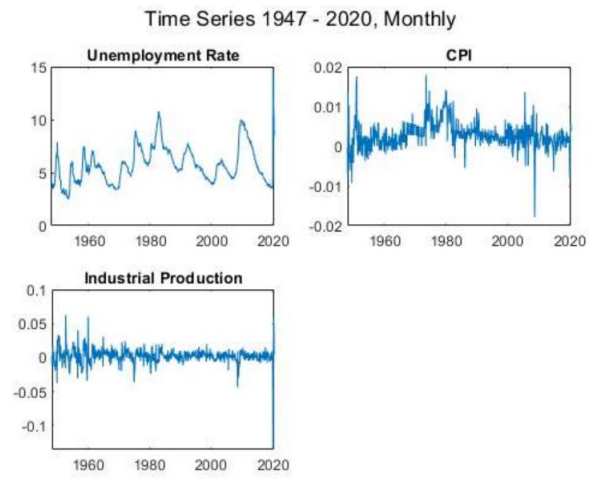
Feel free to drop the number of Monte Carlo simulations to 1000 for computation speed. Remember that when doing Monte-Carlo simulation, each simulation should have its own Newey-West estimator, mean, and data set.

Sometimes you'll receive an error that you cannot Cholesky decompose the Newey-West estimator. Change the number of lags included in the estimator before rewriting your code just to check if that will solve the problem. On a side note, the Newey-West estimator has given me trouble every time I have had to code it so far, so I wish you luck.

Here are some pictures of the histograms and time series I got to guide your journey:









## Chapter 10

# Hypothesis Testing and Confidence Intervals

### 10.1 Previous Problem 1: OLS Estimator

The previous problem set went well, so I will cover a quick question just to review Method of Moments estimation. Looking at question 5 from PS 9:

Consider the  $k \times 1$  vector of population moment conditions

$$\mathbb{E}[\mathbf{x}_i(y_i - \mathbf{x}_i'\boldsymbol{\beta})] = 0$$

where  $\boldsymbol{\beta}$  is a  $k \times 1$  vector of unknown parameters,  $y_i$  is a scalar and  $\mathbf{x}_i$  is a  $k \times 1$  vector.

- Solve for the method of moments population parameter  $\boldsymbol{\beta}_0$ .
- In the solution to (a), the population method of moments parameter  $\boldsymbol{\beta}_0$  is only identified if what condition holds?
- Suppose you observe an *i.i.d.* sample  $\{y_i, \mathbf{x}_i\}_{i=1}^n$  from some unknown "true" joint distribution  $f(y_i, \mathbf{x}_i)$ . Using the analogy principle, propose estimators for the following population moments:

$$\begin{aligned} &\mathbb{E}[\mathbf{x}_i\mathbf{x}_i'] \\ &\mathbb{E}[\mathbf{x}_i\mathbf{x}_i']^{-1} \\ &\mathbb{E}[\mathbf{x}_iy_i] \end{aligned}$$

Are the estimators you proposed consistent for the population moments?

- Consider again the *i.i.d.* sample  $\{y_i, \mathbf{x}_i\}_{i=1}^n$  from  $f(y_i, \mathbf{x}_i)$ . Given your solution to part (a), use the analogy principle to propose a method of moments estimator  $\hat{\boldsymbol{\beta}}_n$  for the population parameter  $\boldsymbol{\beta}_0$ .

### 10.1.1 Part a

We first note the moment conditions:

$$\mathbb{E}[\mathbf{x}_i(y_i - \mathbf{x}'_i\boldsymbol{\beta})] = 0$$

Now, solve for  $\boldsymbol{\beta}_0$ :

$$\begin{aligned}\mathbb{E}[\mathbf{x}_i(y_i - \mathbf{x}'_i\boldsymbol{\beta})] &= 0 \\ \mathbb{E}[\mathbf{x}_iy_i] - \mathbb{E}[\mathbf{x}_i\mathbf{x}'_i]\boldsymbol{\beta} &= 0 \\ \mathbb{E}[\mathbf{x}_iy_i] &= \mathbb{E}[\mathbf{x}_i\mathbf{x}'_i]\boldsymbol{\beta} \\ \mathbb{E}[\mathbf{x}_i\mathbf{x}'_i]^{-1}\mathbb{E}[\mathbf{x}_iy_i] &= \boldsymbol{\beta}_0\end{aligned}$$

### 10.1.2 Part b

For there to be one unique solution for  $\boldsymbol{\beta}_0$ , we need to be able to solve the equation.  $\mathbb{E}[\mathbf{x}_iy_i]$  shouldn't be a problem, but  $\mathbb{E}[\mathbf{x}_i\mathbf{x}'_i]^{-1}$  may cause issues. If  $\mathbb{E}[\mathbf{x}_i\mathbf{x}'_i]$  is singular, then we will not be able to find  $\boldsymbol{\beta}_0$ . For a matrix to be invertible, it needs to be full rank. Therefore, we need  $\mathbb{E}[\mathbf{x}_i\mathbf{x}'_i]$  to be full rank. This requirement is known as the rank condition.

### 10.1.3 Part c

Simply apply the analogy principle:

$$\mathbb{E}[\mathbf{x}_i\mathbf{x}'_i] \longrightarrow \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i\mathbf{x}'_i \tag{10.1}$$

$$\mathbb{E}[\mathbf{x}_i\mathbf{x}'_i]^{-1} \longrightarrow \left( \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i\mathbf{x}'_i \right)^{-1} \tag{10.2}$$

$$\mathbb{E}[\mathbf{x}_iy_i] \longrightarrow \frac{1}{n} \sum_{i=1}^n \mathbf{x}_iy_i \tag{10.3}$$

We know that (1) and (3) converge to the population moment by the WLLN. Because (1) converges by the WLLN and because the inverse function is continuous over the reals, with the exception of zero (because we assume the matrix is invertible, this is not a problem), (2) converges by the CMT.

### 10.1.4 Part d

Simply plug in our estimators from (c):

$$\hat{\boldsymbol{\beta}}_n = \left( \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i\mathbf{x}'_i \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^n \mathbf{x}_iy_i \right)$$

We have just derived the OLS estimator using the Method of Moments. We now know that OLS is both a Maximum Likelihood Estimator and a Method of Moments of estimator.

## 10.2 Hypothesis Testing

From our undergraduate experience, we should be familiar with hypothesis testing. First, we set up two competing hypotheses:

$$H_0 : \theta = \theta_0$$

$$H_A : \theta \neq \theta_0$$

$H_A$  here is set-up as a two sided test. We could also theoretically set up a one-sided test (e.g.  $H_A < \theta_0$ ) but, in economics at least, we want to remain open to results on either side of  $\theta_0$ .

How do we determine whether we have sufficient evidence to reject  $H_0$  (the null hypothesis)? First, we must set the level of the test,  $\alpha$ .  $\alpha$  determines the maximum probability of obtaining a result at least as extreme as the result we observe that we will accept. In other words, we want to be sure that in less than  $\alpha$  of the draws from the null distribution we obtain our result.

Of course, the null distribution matters here in finite distributions. Since calculating the critical value (the value of a distribution in which more extreme values consists of  $\alpha$  of the area under the distribution) for some distributions is difficult, and in the real world we may not know the distribution under the null hypothesis, we turn to asymptotic theory.

Recall the central limit theorem. For *i.i.d.*  $X$ :

$$\frac{\sqrt{n}(\bar{x} - \mu)}{\sigma} \xrightarrow{d} N(0, 1)$$

So let's make use of this! We can find the critical values associated with the values of  $\alpha$  (usually 0.01, 0.05, or 0.10) we choose easily. To find out if our results provide a value beyond the critical values, we can reject the null.

Remembering that we compare our value with a distribution under the null hypothesis, we find the following test statistic:

$$T(\mathbf{y}, \mu_0) = \frac{\sqrt{n}(\bar{x} - \mu_0)}{\sigma}$$

where  $\mu_0$  is the value stated under the null hypothesis. If the null hypothesis is true, the asymptotic distribution is  $\sim N(0, 1)$ . But if the null hypothesis is not true, then we will draw a value from a much different distribution. We compare our t-statistic to the critical value associated with  $\alpha$  (1.96 if  $\alpha = 0.05$ ) and decide whether to reject.

Another way of determining whether to reject our null hypothesis is to construct a confidence interval. The confidence interval gives us a range of values that,  $1-\alpha$  of the time, will contain the true

parameter value. To construct a confidence interval, we calculate:

$$CI_{1-\alpha} = \left[ \bar{x} - cv \cdot \frac{\sigma}{\sqrt{n}}, \quad \bar{x} + cv \cdot \frac{\sigma}{\sqrt{n}} \right]$$

Where  $cv$  denotes the critical value associated with the asymptotic level.

If our null hypothesis is inside the confidence interval, we cannot reject the null at our  $\alpha$  level.

## 10.3 Other Tests

There are other tests that we can do to reject null hypotheses.

### 10.3.1 Likelihood Ratio Test

The likelihood ratio test statistic  $LR(\mathbf{y})$  for testing  $H_0 : \theta \in \Theta_0$  versus  $H_1 : \theta \in \Theta_0^c$  is, in logs:

$$lr(\mathbf{y}) = 2 \left[ \ln(L(\hat{\theta}|\mathbf{y})) - \ln(L(\hat{\theta}_0|\mathbf{y})) \right]$$

where  $L(\cdot)$  denotes the likelihood function from the MLE section.

The asymptotic distribution of the likelihood ratio is:

$$lr(\mathbf{y}) \xrightarrow{d} \chi_r^2$$

where  $r$  is the number of restrictions (null hypotheses tested).

### 10.3.2 Wald Test

The Wald test is as follows:

$$T_w(\mathbf{y}) = n\mathbf{g}(\hat{\theta})' \left( \frac{\partial \mathbf{g}(\hat{\theta})}{\partial \boldsymbol{\theta}} \mathcal{I}_{\hat{\theta}_n, 1}^{-1} \left( \frac{\partial \mathbf{g}(\hat{\theta})}{\partial \boldsymbol{\theta}} \right)' \right)^{-1} \mathbf{g}(\hat{\theta})$$

where  $\mathbf{g}(\boldsymbol{\theta})$  relates the true  $\boldsymbol{\theta}$  and the null  $\boldsymbol{\theta}^0$ . Usually, we like a linear  $\mathbf{g}(\boldsymbol{\theta})$ :

$$\mathbf{g}(\boldsymbol{\theta}) = \boldsymbol{\theta} - \boldsymbol{\theta}_0 = 0$$

as the resulting Wald statistic is the square of the t-test!

The asymptotic distribution of the Wald test is:

$$T_w(\mathbf{y}) \xrightarrow{d} \chi_r^2$$

the same as the Likelihood Ratio test.

### 10.3.3 Score Test

The score test is as follows:

$$T_s(\mathbf{y}) = \frac{1}{n} \frac{\partial \ell_n(\hat{\boldsymbol{\theta}}_n^0)}{\partial \boldsymbol{\theta}'} \mathcal{I}_{\hat{\boldsymbol{\theta}}_{n,1}^0}^{-1} \left( \frac{\partial \ell_n(\hat{\boldsymbol{\theta}}_n^0)}{\partial \boldsymbol{\theta}'} \right)'$$

This too has asymptotic distribution:

$$T_s(\mathbf{y}) \xrightarrow{d} \chi_r^2$$

Note that the superscript indicates whether we are using the null value or our estimate when setting up the tests.

## 10.4 Practice Problem 1: Hansen 13.3 Extended

Take the exponential model with parameter  $\lambda$ . We want a test for  $H_0 : \lambda = 1$  against  $H_A : \lambda \neq 1$ .

- Develop a test based on the sample mean  $\bar{X}_n$ .
- Find the likelihood ratio statistic.
- Find the score test.
- Find the Wald statistic.

### 10.4.1 Part a

We begin with the standard asymptotic t-test:

$$\begin{aligned} T(\mathbf{x}, \mu_0) &= \frac{\sqrt{n}(\bar{x} - \mu_0)}{\hat{s}} \\ &= \frac{\sqrt{n}(\bar{x} - \lambda_0)}{\hat{s}} \\ &= \frac{\sqrt{n}(\bar{x} - 1)}{\hat{s}} \end{aligned}$$

This is our asymptotic t-test.

Note that in this scenario, because we know the true distribution, we can actually calculate the standard deviation:

$$\begin{aligned}
\text{Var}(\sqrt{n}\bar{x}) &= n\text{Var}\left(\frac{1}{n}\sum_{i=1}^n x_i\right) \\
&= \frac{1}{n}\sum_{i=1}^n \text{Var}(x_i) \\
&= \frac{1}{n}\sum_{i=1}^n \hat{\lambda}^2 \\
&= \hat{\lambda}^2 \\
&= \bar{x}^2
\end{aligned}$$

Then our t-test becomes:

$$\begin{aligned}
T(\mathbf{x}) &= \frac{\sqrt{n}(\bar{x} - 1)}{\hat{s}} \\
&= \frac{\sqrt{n}(\bar{x} - 1)}{\sqrt{\bar{x}^2}} \\
&= \frac{\sqrt{n}(\bar{x} - 1)}{\bar{x}}
\end{aligned}$$

### 10.4.2 Part b

Using the formula for the likelihood ratio statistic, and noting that the exponential distribution has a pdf of  $f(x|\lambda) = \frac{1}{\lambda}e^{-\frac{x}{\lambda}}$ :

$$\begin{aligned}
lr(\mathbf{x}) &= 2 \left[ \ln(L(\hat{\theta}|\mathbf{y})) - \ln(L(\hat{\theta}_0|\mathbf{y})) \right] \\
&= 2 \left[ \ln \left( \prod_{i=1}^n \frac{1}{\hat{\lambda}} e^{-\frac{x_i}{\hat{\lambda}}} \right) - \ln \left( \prod_{i=1}^n \frac{1}{1} e^{-\frac{x_i}{1}} \right) \right] \\
&= 2 \left[ \sum_{i=1}^n \left( -\ln(\hat{\lambda}) - \frac{1}{\hat{\lambda}} x_i \right) - \sum_{i=1}^n -x_i \right] \\
&= 2 \left[ -n\ln(\hat{\lambda}) - \frac{n}{\hat{\lambda}} \bar{x}_n + n\bar{x}_n \right] \\
&= 2 \left[ \bar{x}_n \left( n - \frac{n}{\hat{\lambda}} \right) - n\ln(\hat{\lambda}) \right] \\
&\xrightarrow{d} \chi_1^2
\end{aligned}$$



### 10.4.3 Part c

We first need to find the score evaluated at the null hypothesis of  $\lambda_0 = 1$ :

$$\begin{aligned} \left[ \frac{d\ell_n(\lambda)}{d\lambda} \right]_{\lambda=\lambda_0} &= -\frac{n}{\lambda_0} + \frac{n}{\lambda_0^2} \bar{x}_n \\ &= -\frac{n}{1} + \frac{n}{1} \bar{x}_n \\ &= n(\bar{x}_n - 1) \end{aligned}$$

Assuming the information matrix equality holds, the negative Hessian, and thus the information, evaluated at the null value is:

$$\begin{aligned} \mathcal{I}_{\lambda_0,1} &= \frac{1}{\lambda_0^2} \\ &= 1 \end{aligned}$$

So the score test is:

$$\begin{aligned} T_s(\mathbf{x}) &= \frac{1}{n} n(\bar{x}_n - 1) \cdot 1 \cdot n(\bar{x}_n - 1) \\ &= n(\bar{x}_n - 1)^2 \\ &\xrightarrow{d} \chi_1^2 \end{aligned}$$

### 10.4.4 Part d

Last test to derive. Using the constraint  $\mathbf{g}(\cdot)$  defined above:

$$\begin{aligned} \mathbf{g}(\lambda) &= \lambda - \lambda_0 \\ &= \lambda - 1 \\ &= 0 \end{aligned}$$

Now we need the derivative of  $\mathbf{g}(\lambda)$  with respect to the **true**  $\lambda$ :

$$\frac{\partial \mathbf{g}(\lambda)}{\partial \lambda} = 1$$

Using the Wald statistic formula from above:

$$\begin{aligned}
T_w(\mathbf{x}) &= n\mathbf{g}(\hat{\lambda}_n) \left( \frac{\partial \mathbf{g}(\lambda)}{\partial \lambda} \mathcal{I}_{\hat{\lambda}_n, 1}^{-1} \frac{\partial \mathbf{g}(\lambda)}{\partial \lambda} \right)^{-1} \mathbf{g}(\hat{\lambda}_n) \\
&= n(\hat{\lambda}_n - 1) \cdot \frac{1}{\hat{\lambda}_n^2} \cdot (\hat{\lambda}_n - 1) \\
&= \frac{n(\hat{\lambda}_n - 1)^2}{\hat{\lambda}_n^2} \\
&= \frac{n(\bar{x}_n - 1)^2}{\bar{x}_n^2} \\
&\xrightarrow{d} \chi_1^2
\end{aligned}$$

This result is the square of our t-test as expected.

## 10.5 Some Matlab Help

I did not code up your Matlab problem myself this week. Instead, I rely on Drew's sample code from last year. Ultimately, you end up finding the values for the likelihood ratio, Wald, and score tests for a set up data where we assume the true DGP is Poisson. Therefore, your score vectors, Hessians, and information matrices should be derived from this Poisson distribution.

As checkpoints for you while coding, here are the values you should reach for each test if you set your random number generator seed to 10:

```
27.1726  26.8904  26.9001
```

where the order is likelihood ratio first, Wald second, and score third.

Good luck on the homework!

# Chapter 11

## Basics of Regression

### 11.1 Previous Problem 1: Hansen 13.1 Extended

Take the Bernoulli model with probability parameter  $p$ . We want a test for  $H_0 : p = 0.05$  against  $H_1 : p \neq 0.05$ .

- (a) Develop a test based on the sample mean  $\bar{x}_n$ .
- (b) Derive the likelihood ratio statistic. What is its asymptotic sampling distribution?
- (c) Derive the score test. What is its asymptotic sampling distribution?
- (d) Derive the Wald test. What is its asymptotic sampling distribution?

#### 11.1.1 Part a

Because we know that the mean of a Bernoulli random variable is  $p$ , our estimator  $\hat{p}$  is simply  $\bar{x}_n$ . We use a t-test for (a):

$$\begin{aligned} T &= \frac{\sqrt{n}(\bar{x}_n - p_0)}{\hat{s}} \\ &= \frac{\sqrt{n}(\bar{x}_n - 0.05)}{\hat{s}} \end{aligned}$$

Because we know the distribution of our random variable, we replace  $\hat{s}$  with the estimated true variance. Bernoulli distributions have variance  $p(1 - p)$ , so our t-test becomes:

$$\begin{aligned}
T &= \frac{\sqrt{n}(\bar{x}_n - 0.05)}{\hat{s}} \\
&= \frac{\sqrt{n}(\bar{x}_n - 0.05)}{\sqrt{\hat{p}(1 - \hat{p})}} \\
&= \frac{\sqrt{n}(\bar{x}_n - 0.05)}{\sqrt{\bar{x}_n(1 - \bar{x}_n)}}
\end{aligned}$$

### 11.1.2 Part b

Our  $\hat{p}_0 = 0.05$ , so using the formula for the likelihood ratio statistic:

$$\begin{aligned}
lr_n &= 2 \left[ \sum_{i=1}^n \ln(\hat{p}^{y_i}(1 - \hat{p})^{1-y_i}) - \sum_{i=1}^n \ln(\hat{p}_0^{y_i}(1 - \hat{p}_0)^{1-y_i}) \right] \\
&= 2 \left[ \sum_{i=1}^n y_i \ln(\hat{p}) + (1 - y_i) \ln(1 - \hat{p}) - \sum_{i=1}^n y_i \ln(0.05) + (1 - y_i) \ln(1 - 0.05) \right] \\
&\xrightarrow{d} \chi_1^2
\end{aligned}$$

Where the one degree of freedom comes from the one restriction we impose (the null hypothesis).

### 11.1.3 Part c

To derive the score test, we first need to find the score vector evaluated at the restricted value:

$$\begin{aligned}
\left[ \frac{\partial \ell_n(p)}{\partial p} \right]_{p=p_0} &= \left[ \sum_{i=1}^n \frac{y_i}{p} - \sum_{i=1}^n \frac{1 - y_i}{1 - p} \right]_{p=p_0} \\
&= \sum_{i=1}^n \frac{y_i}{0.05} - \sum_{i=1}^n \frac{1 - y_i}{1 - 0.05} \\
&= \frac{n\bar{y}_n}{0.05} - \frac{n - n\bar{y}_n}{1 - 0.05} \\
&= n\bar{y}_n \left[ 20 + \frac{20}{19} \right] - \frac{20n}{19} \\
&= n\bar{y}_n \left[ \frac{400}{19} \right] - \frac{20n}{19}
\end{aligned}$$

We now need the information matrix evaluated under the null hypothesis. Knowing that this parametric model is correctly specified (due to us knowing the underlying distribution) and regular, we can use the information matrix equality:

$$\begin{aligned}
\mathcal{I}_{\hat{p}_0,1} &= \mathbb{E} \left[ -\frac{\partial \ell_n(p)}{\partial p} \right]_{p=p_0} \\
&= \mathbb{E} \left[ \frac{\bar{y}}{p^2} - \frac{\bar{y}}{(1-p)^2} + \frac{1}{(1-p)^2} \right]_{p=p_0} \\
&= \left[ \frac{p}{p^2} - \frac{p}{(1-p)^2} + \frac{1}{(1-p)^2} \right]_{p=p_0} \\
&= \left[ \frac{1}{p} + \frac{1}{1-p} \right]_{p=p_0} \\
&= \left[ \frac{1-p+p}{p(1-p)} \right]_{p=p_0} \\
&= \left[ \frac{1}{p(1-p)} \right]_{p=p_0} \\
&= \frac{1}{0.05(1-0.05)} \\
&= \frac{400}{19}
\end{aligned}$$

We now have all of the pieces that we need. Plugging these into the score statistic formula:

$$\begin{aligned}
T_s &= \frac{1}{n} \left( n\bar{y} \left[ \frac{400}{19} \right] - \frac{20n}{19} \right) \frac{19}{400} \left( n\bar{y} \left[ \frac{400}{19} \right] - \frac{20n}{19} \right) \\
&\xrightarrow{d} \chi_1^2
\end{aligned}$$

#### 11.1.4 Part d

We first need to set up our  $\mathbf{g}$ . As I noted last week, the easiest way to do this is linearly:

$$\begin{aligned}
\mathbf{g}(p) &= p - p_0 \\
&= p - 0.05 \\
&= 0
\end{aligned}$$

Next we need to take the derivative of  $\mathbf{g}(p)$  with respect to  $p$ :

$$\frac{\partial \mathbf{g}(p)}{\partial p} = 1$$

Taking  $\mathcal{I}_{\hat{p}_0,1}$  we estimated above, simply replace  $p$  with  $\hat{p}$  and we have the information matrix we need. Plugging everything into the Wald statistic formula:

$$\begin{aligned}
T_w &= n(\hat{p} - 0.05)(1 \cdot \hat{p}(1 - \hat{p}) \cdot 1)^{-1}(\hat{p} - 0.05) \\
&= \frac{n(\hat{p} - 0.05)^2}{\hat{p}(1 - \hat{p})} \\
&\xrightarrow{d} \chi_1^2
\end{aligned}$$

Notice that because  $\hat{p} = \bar{x}_n$  and because  $\mathbf{g}(p)$  is linear, our Wald statistic is simply the square of our t-statistic.

## 11.2 Regression Theory

Know the basics for linear regression theory. Putting in the work now will pay-off next semester in Metrics II. Some of the material discussed below will seem unimportant now but will become important when proving results in the Spring.

### 11.2.1 Two Special Matrices

#### Projection Matrix

Let  $\mathbf{X}$  be an  $n \times k$  matrix that is full rank. Then the projection matrix  $\mathbf{P}$  is an  $n \times n$  that results from:

$$\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$

The projection matrix can be shown to have the following properties:

- (i)  $\mathbf{PX} = \mathbf{X}$
- (ii)  $\mathbf{P} = \mathbf{P}'$
- (iii)  $\mathbf{PP} = \mathbf{P}$
- (iv)  $tr(\mathbf{P}) = k$  and  $rank(\mathbf{P}) = k$

We can use the projection matrix to find estimated  $\hat{y}$  values in our regressions.

#### Annihilator Matrix

Let  $\mathbf{X}$  be an  $n \times k$  matrix that is full rank. Then the projection matrix  $\mathbf{M}$  is an  $n \times n$  that results from:

$$\begin{aligned}
\mathbf{M} &= \mathbf{I} - \mathbf{P} \\
&= \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'
\end{aligned}$$

The annihilator matrix can be shown to have the following properties:

- (i)  $\mathbf{M}\mathbf{X} = \mathbf{0}$
- (ii)  $\mathbf{M}\mathbf{P} = \mathbf{0}$
- (iii)  $\mathbf{M} = \mathbf{M}'$
- (iv)  $\mathbf{M}\mathbf{M} = \mathbf{M}$
- (v)  $\text{tr}(\mathbf{M}) = n - k$  and  $\text{rank}(\mathbf{M}) = n - k$

We can use the annihilator matrix to remove parts of the regression that we are not interested in estimating. Why you would do this may not be obvious now but you will see its usefulness when you cover the Frisch-Waugh-Lovell Theorem. We can also use the annihilator matrix to find residuals.

### 11.2.2 Normal Linear Regression

The normal linear regression model is as follows:

$$y_i = \mathbf{x}_i' \boldsymbol{\beta} + \varepsilon_i \quad \varepsilon_i \sim N(0, \sigma^2)$$

When using normal linear regression, we assume the following:

- (i) The true model is linear and  $\mathbb{E}[\varepsilon|\mathbf{X}] = 0$ .
- (ii) Data  $\{y_i, \mathbf{x}_i\}$  are *i.i.d.*
- (iii)  $\mathbb{E}[|y_i|] < \infty$  and  $\mathbb{E}[|\mathbf{x}_i|^2] < \infty$ .
- (iv)  $\mathbb{E}[\mathbf{x}_i \mathbf{x}_i']$  is full rank.
- (v) Conditional errors  $\varepsilon_i | \mathbf{x}_i$  are Gaussian.

Using maximum likelihood estimation, since we know the underlying DGP is normal, we can find that our estimators are:

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= (\mathbf{x}'\mathbf{x})^{-1} \mathbf{x}'\mathbf{y} \\ &= \left( \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i' \mathbf{x}_i \right) \left( \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i y_i \right) \\ \hat{\sigma}^2 &= \frac{1}{n} \sum_{i=1}^n (y_i - \mathbf{x}_i' \hat{\boldsymbol{\beta}})^2 \\ &= \frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_i^2 \end{aligned}$$

Using normal linear regression, we can easily derive finite sampling distributions. Usually we don't want to assume normally distributed errors. Luckily for us, the best linear predictor follows the same process! But we will need to appeal to asymptotic theory to conduct our hypothesis testing.

### 11.2.3 Assumptions on the Error Terms

I will lay out the two main error assumptions here.

#### Strict Exogeneity

Strict exogeneity is the strongest assumption we can impose on the data. We have strict exogeneity when:

$$\mathbb{E}[\varepsilon_i | \mathbf{x}_i] = 0 \quad \forall i$$

Strict exogeneity is nice as it implies that the errors have a mean of zero:

$$\begin{aligned} \mathbb{E}[\varepsilon_i] &= \mathbb{E}_{\mathbf{x}} [\mathbb{E}[\varepsilon_i | \mathbf{x}_i]] && \text{Using LIE} \\ &= \mathbb{E}_{\mathbf{x}} [0] \\ &= 0 \end{aligned}$$

In addition, strict exogeneity implies that the covariance between  $\mathbf{x}_i$  and  $\varepsilon_i$  is zero:

$$\begin{aligned} \text{Cov}(\mathbf{x}_i, \varepsilon_i) &= \mathbb{E}[\mathbf{x}_i \varepsilon_i] - \mathbb{E}[\mathbf{x}_i] \mathbb{E}[\varepsilon_i] \\ &= \mathbb{E}[\mathbf{x}_i \varepsilon_i] \\ &= \mathbb{E}_{\mathbf{x}} \mathbb{E}[\mathbf{x}_i \varepsilon_i | \mathbf{x}_i] \\ &= \mathbb{E}_{\mathbf{x}} [\mathbf{x}_i \mathbb{E}[\varepsilon_i | \mathbf{x}_i]] && \text{Due to the conditional} \\ &= \mathbb{E}_{\mathbf{x}} [\mathbf{x}_i \cdot 0] \\ &= 0 \end{aligned}$$

So strict exogeneity provides lots of nice properties. But this is a stronger assumption than we need.

#### Orthogonality (Exogenous Variables)

We actually only need the orthogonality condition:

$$\mathbb{E}[\mathbf{x}_i \varepsilon_i] = \mathbf{0}$$

Note that we proved that this result follows naturally from the covariance proof for strict exogeneity. As such, strict exogeneity implies orthogonality.



### 11.2.4 The OLS Estimator

Sometimes our regression doesn't have normal errors or a conditional expectation function underlying the model that is linear. As such, if we still want to use our regression tools, we need to use a best linear predictor. We use the OLS estimator for this job. Note that if the CEF is non-linear, OLS is the best *linear* predictor for the CEF but is not the best predictor for the CEF.

As stated above, our OLS estimator under a non-normal regression is the same as the maximum likelihood estimator for a normal regression. To see this, we start with minimizing the sum of squared residuals:

$$\begin{aligned}\beta &= \underset{\beta \in \mathbb{R}^k}{\operatorname{argmin}} \mathbb{E} [(y_i - \mathbf{x}'_i \beta)'(y_i - \mathbf{x}'_i \beta)] \\ &= \mathbb{E} [-\mathbf{x}_i(y_i - \mathbf{x}'_i \beta) - \mathbf{x}_i(y_i - \mathbf{x}'_i \beta)] \\ &= \mathbb{E} [-\mathbf{x}'_i y_i + \mathbf{x}'_i \mathbf{x}_i \beta - \mathbf{x}'_i y_i + \mathbf{x}'_i \mathbf{x}_i \beta]\end{aligned}$$

Because we are minimizing, we set this equal to zero:

$$\begin{aligned}0 &= 2\mathbb{E}[\mathbf{x}'_i \mathbf{x}_i \beta] - 2\mathbb{E}[\mathbf{x}'_i y_i] \\ \mathbb{E}[\mathbf{x}'_i \mathbf{x}_i] \beta &= \mathbb{E}[\mathbf{x}'_i y_i] \\ \beta &= \mathbb{E}[\mathbf{x}'_i \mathbf{x}_i]^{-1} \mathbb{E}[\mathbf{x}_i y_i]\end{aligned}$$

To estimate, we employ the analogy principle:

$$\hat{\beta}_{OLS} = \left( \frac{1}{n} \sum_{i=1}^n \mathbf{x}'_i \mathbf{x}_i \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^n \mathbf{x}'_i y_i \right)$$

### 11.2.5 Consistency and Asymptotic Normality

For most estimators, we want both consistency for obvious reasons and asymptotic normality for hypothesis testing.

#### Consistency

If we are going to use OLS, we should make sure that under our assumptions OLS is consistent for the best linear predictor. To do this, we start from the estimator derived above and proceed to simplify until we can use our asymptotic theorems: WLLN, CMT, and Slutsky's. Completing the proof requires us to recognize the fact that  $y_i = \mathbf{x}_i \beta + \varepsilon_i$ :

$$\begin{aligned}
\hat{\beta}_{OLS} &= \left( \frac{1}{n} \sum_{i=1}^n \mathbf{x}'_i \mathbf{x}_i \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^n \mathbf{x}'_i y_i \right) \\
&= \left( \frac{1}{n} \sum_{i=1}^n \mathbf{x}'_i \mathbf{x}_i \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^n \mathbf{x}'_i (\mathbf{x}_i \beta + \varepsilon_i) \right) \\
&= \left( \frac{1}{n} \sum_{i=1}^n \mathbf{x}'_i \mathbf{x}_i \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^n \mathbf{x}'_i \mathbf{x}_i \beta \right) + \left( \frac{1}{n} \sum_{i=1}^n \mathbf{x}'_i \mathbf{x}_i \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^n \mathbf{x}'_i \varepsilon_i \right) \\
&= \beta + \left( \frac{1}{n} \sum_{i=1}^n \mathbf{x}'_i \mathbf{x}_i \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^n \mathbf{x}'_i \varepsilon_i \right)
\end{aligned}$$

Taking the limit as  $n \rightarrow \infty$ , we can apply the WLLN and CMT to find that:

$$\hat{\beta}_{OLS} \xrightarrow{P} \beta + \mathbb{E}[\mathbf{x}'_i \mathbf{x}_i]^{-1} \mathbb{E}[\mathbf{x}'_i \varepsilon_i]$$

Recalling our assumption of orthogonality, we know that this becomes:

$$\begin{aligned}
&= \beta + \mathbb{E}[\mathbf{x}'_i \mathbf{x}_i]^{-1} \cdot \mathbf{0} \\
&= \beta
\end{aligned}$$

Therefore:

$$\hat{\beta}_{OLS} \xrightarrow{P} \beta$$

### Asymptotic Normality

So our OLS estimate is consistent. Now we want to ensure that we can easily test hypotheses. So we want to see if it is asymptotically normal. We follow the same set-up steps as we did when proving consistency:

$$\begin{aligned}
\hat{\beta}_{OLS} &= \left( \frac{1}{n} \sum_{i=1}^n \mathbf{x}'_i \mathbf{x}_i \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^n \mathbf{x}'_i y_i \right) \\
&= \left( \frac{1}{n} \sum_{i=1}^n \mathbf{x}'_i \mathbf{x}_i \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^n \mathbf{x}'_i (\mathbf{x}_i \beta + \varepsilon_i) \right)
\end{aligned}$$

$$\begin{aligned}
&= \left( \frac{1}{n} \sum_{i=1}^n \mathbf{x}'_i \mathbf{x}_i \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^n \mathbf{x}'_i \mathbf{x}_i \boldsymbol{\beta} \right) + \left( \frac{1}{n} \sum_{i=1}^n \mathbf{x}'_i \mathbf{x}_i \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^n \mathbf{x}'_i \varepsilon_i \right) \\
&= \boldsymbol{\beta} + \left( \frac{1}{n} \sum_{i=1}^n \mathbf{x}'_i \mathbf{x}_i \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^n \mathbf{x}'_i \varepsilon_i \right)
\end{aligned}$$

Now, we subtract  $\boldsymbol{\beta}$  to the left-hand side and multiply both sides by  $\sqrt{n}$ :

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{OLS} - \boldsymbol{\beta}) = \left( \frac{1}{n} \sum_{i=1}^n \mathbf{x}'_i \mathbf{x}_i \right)^{-1} \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{x}'_i \varepsilon_i \right)$$

This set-up should look familiar. The left-hand side now looks like the Central Limit Theorem. We now take advantage of the Multivariate CLT, which says that if we have:

$$\begin{aligned}
\frac{1}{n} \sum_{i=1}^n \omega_i & \quad \mathbb{E}[\omega_i] = 0 \quad \text{Var}(\omega_i) = \Sigma \\
\text{then} & \quad \sqrt{n} \left( \frac{1}{n} \sum_{i=1}^n \omega_i \right) \xrightarrow{d} N(0, \Sigma).
\end{aligned}$$

Note that  $\frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{x}'_i \varepsilon_i$  meets each of these requirements, so if we take the limit as  $n \rightarrow \infty$ :

$$\begin{aligned}
\sqrt{n}(\hat{\boldsymbol{\beta}}_{OLS} - \boldsymbol{\beta}) & \xrightarrow{d} \mathbb{E}[\mathbf{x}'_i \mathbf{x}_i]^{-1} \cdot N(0, \mathbb{E}[\mathbf{x}'_i \varepsilon_i \varepsilon'_i \mathbf{x}_i]) \\
& = N(0, \mathbb{E}[\mathbf{x}'_i \mathbf{x}_i]^{-1} \mathbb{E}[\mathbf{x}'_i \varepsilon_i \varepsilon'_i \mathbf{x}_i] \mathbb{E}[\mathbf{x}'_i \mathbf{x}_i]^{-1})
\end{aligned}$$

The asymptotic variance above is the robust asymptotic variance (this simply means that we do not assume homoskedasticity in the error terms:  $\text{Var}(\varepsilon_i | \mathbf{x}_i) = \mathbb{E}[\varepsilon_i^2 | \mathbf{x}_i] = \sigma^2(\mathbf{x}_i)$ ).

If we do assume homoskedasticity ( $\text{Var}(\varepsilon_i | \mathbf{x}_i) = \mathbb{E}[\varepsilon_i^2 | \mathbf{x}_i] = \sigma^2(\mathbf{x}_i) = \sigma^2 \forall \mathbf{x}_i$ ), then we converge to:

$$\begin{aligned}
\sqrt{n}(\hat{\boldsymbol{\beta}}_{OLS} - \boldsymbol{\beta}) & \xrightarrow{d} N(0, \sigma^2 \mathbb{E}[\mathbf{x}'_i \mathbf{x}_i]^{-1} \mathbb{E}[\mathbf{x}'_i \mathbf{x}_i] \mathbb{E}[\mathbf{x}'_i \mathbf{x}_i]^{-1}) \\
& = N(0, \sigma^2 \mathbb{E}[\mathbf{x}'_i \mathbf{x}_i]^{-1})
\end{aligned}$$

Know these proofs backward and forward and know how the assumptions we make play into the proof. Having a strong foundation built for consistency and asymptotic normality proof will be extremely helpful for Metrics II.

## 11.3 Matlab Help

Today's recitation was full of theory, as there was lots to go through. I wanted to get through a big chunk of regression theory as I am not sure what the next recitation session will hold (depending on scheduling I may use the next recitation to prepare for the final). Even still, I think providing some Matlab help is prudent.

For this next problem set, you get to recreate Stata output in Matlab. This Matlab exercise is one of the most useful - and cool - problems you will do this semester. To make both this exercise and, potentially, future exercises easier, create your regression function in a separate file. Create this function in such a way that it outputs the tables with labels automatically.

In addition to the homoskedastic standard errors, try also including robust standard errors in your table and compare the numbers. To make sure your output is correct, you can always input the data into Stata and compare.

Note that to include a constant in your regression, add a vector of ones as a new "variable" to your data. Here is a snapshot of the inputs I have to my function:

```
o = ones(length(data)-1,1);
x0 = data(2:end, [1,3,4,5]);
x = [x0,o];
y = data(2:end,2);
a = .95;
n = 114;
k = 5;
variables = ["Bedrooms"; "Bathrooms"; "Other Rooms"; "Age"; "Constant"];
names = ["Variable", "Coef.", "Std. Err.", "Robust Std. Err.",...
        "t", "P value", "Conf. Int. (lower)", "Conf. Int. (upper)"];
info = ["Num Obs"; strcat("F[" ,string(k-1),",",string(n-k),"]"); ...
        "Prob > F"; "R-Squared"; "Adj R-Squared"; "Root MSE"];
headings = ["Source", 'SS', 'df', "MS"];
labels = ["Model"; "Residual"; "Total"];

robustregress(x,y,n,k,a, headings, labels, info, variables, names)
```

Here is the code I use to generate my tables:

```
% Table
data = [n; ftest; fvalue; r2; adjr2; mse];
table_1 = table(labels,[ssm; ssr; tss], [k-1;n-k;n-1], [msm; msr;tms], 'VariableNames',headings)

table_2 = table(info, data)

Regression_table = table(variables, beta1, stderhom, stderhet, t', p', cilow', cihigh', 'VariableNames',names)
```

And here is how my table output looks like for Model 2 (the confidence intervals still use homoskedastic standard errors even though I also report robust standard errors):

```

table_1 =
3x4 table
  Source      SS      df      MS
-----
"Model"      1.6042e+06  5      3.2085e+05
"Residual"   2.4241e+06 108     22445
"Total"      4.0283e+06 113     35649

table_2 =
6x2 table
  info      data
-----
"Num Obs"   114
"F[4,109]"  14.295
"Prob > F"  9.9213e-11
"R-Squared" 0.39824
"Adj R-Squared" 0.37038
"Root MSE"  149.82

Regression_table =
6x8 table
  Variable      Coef.      Std. Err.      Robust Std. Err.      t      P value      Conf. Int. (lower)      Conf. Int. (upper)
-----
"Bedrooms"     -21.915     18.394           27.575           -1.1914     0.23612           -58.376           14.546
"Bathrooms"    -0.96385     32.174           32.741           -0.029958    0.97616           -64.738           62.81
"Other Rooms"  -5.3018     14.031           15.501           -0.37788    0.70626           -33.113           22.509
"Age"          -0.13753     0.44989          0.39923           -0.30571    0.76042           -1.0293           0.75422
"Square Feet"  0.20277     0.037337         0.075133          2.6915      0.00807           0.12876           0.27678
"Constant"     80.739      66.589           86.457           0.93398    0.34797           -51.252           212.73

```

## 11.4 Previous Problem 2: Hansen II 3.13

Let  $\mathbf{D}_1$  and  $\mathbf{D}_2$  be vectors of ones and zeroes, with the  $i^{\text{th}}$  element of  $\mathbf{D}_1$  equaling one if that observation is male and zero if that observation is female ( $\mathbf{D}_2$  being the opposite). Then:

- (a) In the OLS regression

$$\mathbf{Y} = \mathbf{D}_1 \hat{\gamma}_1 + \mathbf{D}_2 \hat{\gamma}_2 + \hat{\boldsymbol{\mu}}$$

show that  $\hat{\gamma}_1$  is the sample mean of the dependent variable among men in the sample and that  $\hat{\gamma}_2$  is the sample mean among women.

- (b) Let  $\mathbf{X}(n \times k)$  be an additional matrix of regressors. Describe in words the transformations

$$\begin{aligned} \mathbf{Y}^* &= \mathbf{Y} - \mathbf{D}_1 \bar{Y}_1 - \mathbf{D}_2 \bar{Y}_2 \\ \mathbf{X}^* &= \mathbf{X} - \mathbf{D}_1 \bar{X}_1' - \mathbf{D}_2 \bar{X}_2' \end{aligned}$$

Where  $\bar{X}_1$  and  $\bar{X}_2$  are the  $k \times 1$  means of the regressors for men and women, respectively.

- (c) Compare  $\tilde{\boldsymbol{\beta}}$  from the OLS regression

$$\mathbf{Y}^* = \mathbf{X}^* \tilde{\boldsymbol{\beta}} + \tilde{\boldsymbol{\epsilon}}$$

with the  $\hat{\boldsymbol{\beta}}$  from the OLS regression

$$\mathbf{Y} = \mathbf{D}_1 \hat{\alpha}_1 + \mathbf{D}_2 \hat{\alpha}_2 + \mathbf{X} \hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\epsilon}}$$

### 11.4.1 Part a

We first take the general formula for the OLS estimator:

$$\beta = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{Y})$$

Looking at the OLS equation we are estimating, we see that

$$\mathbf{X} = \begin{bmatrix} \mathbf{D}_1 & \mathbf{D}_2 \end{bmatrix}$$

Calculating  $\mathbf{X}'\mathbf{X}$  then:

$$\begin{aligned} \mathbf{X}'\mathbf{X} &= \begin{bmatrix} \mathbf{D}'_1\mathbf{D}_1 & \mathbf{D}'_1\mathbf{D}_2 \\ \mathbf{D}'_2\mathbf{D}_1 & \mathbf{D}'_2\mathbf{D}_2 \end{bmatrix} \\ &= \begin{bmatrix} n_1 & 0 \\ 0 & n_2 \end{bmatrix} \end{aligned}$$

We can then invert this by using properties of a diagonal block matrix:

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} \frac{1}{n_1} & 0 \\ 0 & \frac{1}{n_2} \end{bmatrix}$$

Now we need to find the second part of the OLS estimator:

$$\begin{aligned} \mathbf{X}'\mathbf{Y} &= \begin{bmatrix} \mathbf{D}'_1\mathbf{Y} \\ \mathbf{D}'_2\mathbf{Y} \end{bmatrix} \\ &= \begin{bmatrix} \sum_{i=1}^n y_i \mathbb{1}(d_{1,i} = 1) \\ \sum_{i=1}^n y_i \mathbb{1}(d_{2,i} = 1) \end{bmatrix} \end{aligned}$$

Combining the two pieces together, we get:

$$\begin{aligned} \hat{\beta}_{OLS} &= \begin{bmatrix} \frac{1}{n_1} & 0 \\ 0 & \frac{1}{n_2} \end{bmatrix} \begin{bmatrix} \sum_{i=1}^n y_i \mathbb{1}(d_{1,i} = 1) \\ \sum_{i=1}^n y_i \mathbb{1}(d_{2,i} = 1) \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{n_1} \sum_{i=1}^n y_i \mathbb{1}(d_{1,i} = 1) \\ \frac{1}{n_2} \sum_{i=1}^n y_i \mathbb{1}(d_{2,i} = 1) \end{bmatrix} \\ &= \begin{bmatrix} \bar{Y}_1 \\ \bar{Y}_2 \end{bmatrix} \end{aligned}$$

### 11.4.2 Part b

The first transformation demeans  $\mathbf{y}$  so that  $\mathbf{y}^*$  has a sample mean of zero. The second transformation demeans  $\mathbf{X}$  so that  $\mathbf{X}^*$  has a sample mean of zero. When running regressions with these variables, the  $\beta$  coefficients are now deviations from the mean of the data.

### 11.4.3 Part c

Let's start from the second OLS regression. We can rewrite this equation as

$$\begin{aligned} \mathbf{Y} &= \mathbf{X}_1 \hat{\beta}_1 + \mathbf{X}_2 \hat{\beta}_2 + \hat{\mathbf{e}} \\ \mathbf{X}_1 &= \begin{bmatrix} \mathbf{D}_1 & \mathbf{D}_2 \end{bmatrix} \\ \hat{\beta}_1 &= \begin{bmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \end{bmatrix} \\ \mathbf{X}_2 &= \mathbf{X} \\ \hat{\beta}_2 &= \hat{\beta} \end{aligned}$$

Now, we don't care about  $\mathbf{X}_1$ , so we can use the elimination matrix to simplify the problem. Define  $\mathbf{M}_1 \equiv \mathbf{I} - \mathbf{X}'_1(\mathbf{X}'_1\mathbf{X}_1)^{-1}\mathbf{X}_1$ . We can now write  $\hat{\beta}_2$  as:

$$\hat{\beta}_2 = ((\mathbf{M}_1\mathbf{X})'\mathbf{M}_1\mathbf{X})^{-1}(\mathbf{M}_1\mathbf{X})'\mathbf{M}_1\mathbf{Y}$$

Okay, let's look at this equation piece-by-piece. First, we look at  $\mathbf{M}_1\mathbf{Y}$ :

$$\begin{aligned} \mathbf{M}_1\mathbf{Y} &= (\mathbf{I} - \mathbf{X}'_1(\mathbf{X}'_1\mathbf{X}_1)^{-1}\mathbf{X}_1)\mathbf{Y} \\ &= \mathbf{Y} - \mathbf{X}'_1(\mathbf{X}'_1\mathbf{X}_1)^{-1}\mathbf{X}_1\mathbf{Y} \\ &= \mathbf{Y} - \mathbf{X}'_1 \begin{bmatrix} \bar{\mathbf{Y}}_1 \\ \bar{\mathbf{Y}}_2 \end{bmatrix} \\ &= \mathbf{Y} - \mathbf{D}_1\bar{\mathbf{Y}}_1 - \mathbf{D}_2\bar{\mathbf{Y}}_2 \\ &= \mathbf{Y}^* \end{aligned}$$

Now we can look at  $\mathbf{M}_1\mathbf{X}$ :

$$\begin{aligned} \mathbf{M}_1\mathbf{X} &= (\mathbf{I} - \mathbf{X}'_1(\mathbf{X}'_1\mathbf{X}_1)^{-1}\mathbf{X}_1)\mathbf{X} \\ &= \mathbf{X} - \mathbf{X}'_1(\mathbf{X}'_1\mathbf{X}_1)^{-1}\mathbf{X}_1\mathbf{X} \end{aligned}$$

$$\begin{aligned}
&= \mathbf{X} - \mathbf{X}'_1 \begin{bmatrix} \frac{1}{n_1} & 0 \\ 0 & \frac{1}{n_2} \end{bmatrix} \begin{bmatrix} \sum_{i=1}^n x_i \mathbb{1}(d_{1,i} = 1) \\ \sum_{i=1}^n x_i \mathbb{1}(d_{2,i} = 1) \end{bmatrix} \\
&= \mathbf{X} - \mathbf{X}'_1 \begin{bmatrix} \bar{\mathbf{x}}'_1 \\ \bar{\mathbf{x}}'_2 \end{bmatrix} \\
&= \mathbf{X} - \mathbf{D}_1 \bar{\mathbf{x}}'_1 - \mathbf{D}_2 \bar{\mathbf{x}}'_2 \\
&= \mathbf{X}^*
\end{aligned}$$

Now we have all the pieces we need to find  $\hat{\beta}_2$ :

$$\begin{aligned}
\hat{\beta}_2 &= ((\mathbf{M}_1 \mathbf{X})' \mathbf{M}_1 \mathbf{X})^{-1} (\mathbf{M}_1 \mathbf{X})' \mathbf{M}_1 \mathbf{Y} \\
&= ((\mathbf{X}^*)' \mathbf{X}^*)^{-1} ((\mathbf{X}^*)' \mathbf{Y}^*)
\end{aligned}$$

Turning to the first OLS regression, we apply our normal formula for the OLS estimator:

$$\tilde{\beta} = ((\mathbf{X}^*)' \mathbf{X}^*)^{-1} ((\mathbf{X}^*)' \mathbf{Y}^*)$$

which is the same estimator we derived from the second OLS regression. Therefore, the two regressions deliver the same results for the target  $\beta$ 's.

## 11.5 Practice Problem 1: Hansen II 4.16 Adapted

Take the linear homoskedastic CEF:

$$\begin{aligned}
Y^* &= X' \beta + e \\
\mathbb{E}[e|X] &= 0 \\
\mathbb{E}[e^2|X] &= \sigma^2
\end{aligned}$$

and suppose that  $Y^*$  is measured with error. Instead of  $Y^*$ , we observe  $Y = Y^* + u$  where  $u$  is measurement error. Suppose that  $e$  and  $u$  are independent and

$$\begin{aligned}
\mathbb{E}[u|X] &= 0 \\
\mathbb{E}[u^2|X] &= \sigma_u^2(X)
\end{aligned}$$

- Derive an equation for  $Y$  as a function of  $X$ .
- Describe the effect of this measurement error on OLS estimation of  $\beta$  in the feasible regression of the observed  $Y$  on  $X$ .
- Describe the effect (if any) of the measurement error on the asymptotic variance for  $\hat{\beta}$ .



### 11.5.1 Part a

Plug in for  $Y^*$  in the main OLS equation:

$$\begin{aligned} Y^* &= X'\beta + e \\ Y - u &= X'\beta + e \\ Y &= X'\beta + e + u && \text{Let } \eta \equiv e + u \\ &= X'\beta + \eta \end{aligned}$$

### 11.5.2 Part b

Using the standard estimator for  $\beta_{OLS}$ :

$$\begin{aligned} \hat{\beta} &= \left( \frac{1}{n} \sum_{i=1}^n x_i x_i' \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^n x_i y_i \right) \\ &= \left( \frac{1}{n} \sum_{i=1}^n x_i x_i' \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^n x_i (x_i' \beta + \eta_i) \right) \\ &= \left( \frac{1}{n} \sum_{i=1}^n x_i x_i' \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^n x_i x_i' \right) \beta + \left( \frac{1}{n} \sum_{i=1}^n x_i x_i' \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^n x_i \eta_i \right) \\ &= \beta + \left( \frac{1}{n} \sum_{i=1}^n x_i x_i' \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^n x_i (e_i + u_i) \right) \\ &= \beta + \left( \frac{1}{n} \sum_{i=1}^n x_i x_i' \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^n x_i e_i \right) + \left( \frac{1}{n} \sum_{i=1}^n x_i x_i' \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^n x_i u_i \right) \quad (1) \end{aligned}$$

Noting our assumptions that  $\mathbb{E}[e|X] = 0$  and  $\mathbb{E}[u|X] = 0$ , we can derive:

$$\begin{aligned} \mathbb{E}[xe] &= \mathbb{E}_x [x\mathbb{E}[e|X]] && \mathbb{E}[xu] = \mathbb{E}_x [x\mathbb{E}[u|X]] \\ &= \mathbb{E}_x [x \cdot 0] && = \mathbb{E}_x [x \cdot 0] \\ &= \mathbb{E}_x [0] && = \mathbb{E}_x [0] \\ &= 0 && = 0 \end{aligned}$$

Taking these two results and taking the probability limit of  $\hat{\beta}$  gives us:

$$\begin{aligned} \hat{\beta} &\xrightarrow{P} \beta + \mathbb{E}[x_i x_i']^{-1} \mathbb{E}[x_i e_i] + \mathbb{E}[x_i x_i']^{-1} \mathbb{E}[x_i u_i] \\ &= \beta \end{aligned}$$

So we see that  $\hat{\beta}$  is still consistent for  $\beta_{OLS}$ .

### 11.5.3 Part c

Starting from (1) in Part b:

$$\begin{aligned}\hat{\beta} &= \beta + \left(\frac{1}{n} \sum_{i=1}^n x_i x_i'\right)^{-1} \left(\frac{1}{n} \sum_{i=1}^n x_i e_i\right) + \left(\frac{1}{n} \sum_{i=1}^n x_i x_i'\right)^{-1} \left(\frac{1}{n} \sum_{i=1}^n x_i u_i\right) \\ \hat{\beta} - \beta &= \left(\frac{1}{n} \sum_{i=1}^n x_i x_i'\right)^{-1} \left(\frac{1}{n} \sum_{i=1}^n x_i e_i\right) + \left(\frac{1}{n} \sum_{i=1}^n x_i x_i'\right)^{-1} \left(\frac{1}{n} \sum_{i=1}^n x_i u_i\right) \\ \sqrt{n}(\hat{\beta} - \beta) &= \left(\frac{1}{n} \sum_{i=1}^n x_i x_i'\right)^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n x_i e_i\right) + \left(\frac{1}{n} \sum_{i=1}^n x_i x_i'\right)^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n x_i u_i\right)\end{aligned}$$

To simplify the resulting expression, define  $Q_{xx} \equiv \mathbb{E}[x_i x_i']$ . Then as we take  $n \rightarrow \infty$ :

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, Q_{xx}^{-1} \mathbb{E}[x_i e_i e_i' x_i'] Q_{xx}^{-1}) + N(0, Q_{xx}^{-1} \mathbb{E}[x_i u_i u_i' x_i'] Q_{xx}^{-1})$$

Note our assumption that  $\mathbb{E}[e^2|X] = \sigma^2$ . Then:

$$\begin{aligned}\mathbb{E}[e_i^2] &= \mathbb{E}_{x_i} [\mathbb{E}[e_i^2|x_i]] \\ &= \mathbb{E}_{x_i} [\sigma^2] \\ &= \sigma^2\end{aligned}$$

We can simplify our expression for the asymptotic distribution of  $\hat{\beta}$ :

$$\begin{aligned}\sqrt{n}(\hat{\beta} - \beta) &\xrightarrow{d} N(0, \sigma^2 Q_{xx}^{-1}) + N(0, Q_{xx}^{-1} \mathbb{E}[x_i u_i u_i' x_i'] Q_{xx}^{-1}) \\ \sqrt{n}(\hat{\beta} - \beta) &\xrightarrow{d} N(0, \sigma^2 Q_{xx}^{-1} + Q_{xx}^{-1} \mathbb{E}[x_i u_i u_i' x_i'] Q_{xx}^{-1})\end{aligned}$$

Note that we cannot simplify the second term as the conditional measurement errors are heteroskedastic.

## 11.6 Practice Problem 2: Hansen II 7.1

Take the model  $Y = X_1' \beta_1 + X_2' \beta_2 + e$  with  $\mathbb{E}[Xe] = 0$ . Suppose that  $\beta_1$  is estimated regressing  $Y$  on  $X_1$  only. Find the probability limit of the estimator. In general, is it consistent for  $\beta_1$ ? If not, under what conditions is this estimator consistent for  $\beta_1$ ?

### 11.6.1 Solution

We start by defining  $\eta \equiv X_2' \beta_2 + e$ . Then we have:

$$Y = X_1' \beta_1 + \eta$$

We now solve from the  $\hat{\beta}_{OLS}$  estimator:

$$\begin{aligned} \hat{\beta}_1 &= \left( \frac{1}{n} \sum_{i=1}^n x_{1i} x_{1i}' \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^n x_{1i} y_i \right) \\ &= \left( \frac{1}{n} \sum_{i=1}^n x_{1i} x_{1i}' \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^n x_{1i} (x_{1i}' \beta_1 + \eta) \right) \\ &= \left( \frac{1}{n} \sum_{i=1}^n x_{1i} x_{1i}' \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^n x_{1i} x_{1i}' \right) \beta_1 + \left( \frac{1}{n} \sum_{i=1}^n x_{1i} x_{1i}' \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^n x_{1i}' \eta \right) \\ &= \beta_1 + \left( \frac{1}{n} \sum_{i=1}^n x_{1i} x_{1i}' \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^n x_{1i}' (x_{2i}' \beta_2 + e_i) \right) \\ &= \beta_1 + \left( \frac{1}{n} \sum_{i=1}^n x_{1i} x_{1i}' \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^n x_{1i} x_{2i}' \right) \beta_2 + \left( \frac{1}{n} \sum_{i=1}^n x_{1i} x_{1i}' \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^n x_{1i} e_i \right) \\ &\xrightarrow{P} \beta_1 + \mathbb{E}[x_{1i} x_{1i}']^{-1} \mathbb{E}[x_{1i} x_{2i}'] \beta_2 + \mathbb{E}[x_{1i} x_{1i}']^{-1} \mathbb{E}[x_{1i} e_i] \\ &= \beta_1 + \mathbb{E}[x_{1i} x_{1i}']^{-1} \mathbb{E}[x_{1i} x_{2i}'] \beta_2 \end{aligned}$$

If we assume that  $\beta_2 = 0$ , then we see that  $\hat{\beta}_1 \xrightarrow{P} \beta_1$ . But if  $\beta_2 \neq 0$ , then  $\hat{\beta}_1$  does not converge to  $\beta_1$ .

## 11.7 Practice Problem 3: Hansen II 7.15

Take the linear model  $Y = X\beta + e$  with  $\mathbb{E}[e|X] = 0$  and  $X \in \mathbb{R}$ . Consider the estimator

$$\hat{\beta} = \frac{\sum_{i=1}^n x_i^3 y_i}{\sum_{i=1}^n x_i^4}$$

Find the asymptotic distribution of  $\sqrt{n}(\hat{\beta} - \beta)$  as  $n \rightarrow \infty$ .

### 11.7.1 Solution

We take  $\hat{\beta}$  and work from there:

$$\begin{aligned}
\hat{\beta} &= \left( \frac{1}{n} \sum_{i=1}^n x_i^4 \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^n x_i^3 y_i \right) \\
&= \left( \frac{1}{n} \sum_{i=1}^n x_i^4 \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^n x_i^3 (x_i \beta + e_i) \right) \\
&= \left( \frac{1}{n} \sum_{i=1}^n x_i^4 \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^n x_i^4 \right) \beta + \left( \frac{1}{n} \sum_{i=1}^n x_i^4 \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^n x_i^3 e_i \right) \\
&= \beta + \left( \frac{1}{n} \sum_{i=1}^n x_i^4 \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^n x_i^3 e_i \right) \\
\hat{\beta} - \beta &= \left( \frac{1}{n} \sum_{i=1}^n x_i^4 \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^n x_i^3 (x_i e_i) \right)
\end{aligned}$$

Note that the last term is zero in expectation. Therefore, if we multiply both sides by  $\sqrt{n}$  we can apply the multivariate central limit theorem:

$$\begin{aligned}
\sqrt{n}(\hat{\beta} - \beta) &= \left( \frac{1}{n} \sum_{i=1}^n x_i^4 \right)^{-1} \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n x_i^3 (x_i e_i) \right) \\
&\xrightarrow{d} \mathbb{E}[x_i^4]^{-1} N(0, \mathbb{E}[x_i^3 e_i x_i^3]) \\
&= N(0, \mathbb{E}[x_i^4]^{-1} \mathbb{E}[x_i^3 e_i x_i^3] \mathbb{E}[x_i^4]^{-1})
\end{aligned}$$

## 11.8 Practice Problem 4: Hansen II 7.23 Adapted

The model is  $Y = X\beta + e$  with  $\mathbb{E}[e|X] = 0$  and  $X \in \mathbb{R}$ . Consider the estimator:

$$\tilde{\beta} = \frac{1}{n} \sum_{i=1}^n \frac{y_i}{x_i}$$

Is  $\tilde{\beta}$  consistent for  $\beta$  as  $n \rightarrow \infty$ ?

### 11.8.1 Solution

Starting from  $\tilde{\beta}$ :

$$\begin{aligned}
\tilde{\beta} &= \frac{1}{n} \sum_{i=1}^n \frac{y_i}{x_i} \\
&= \frac{1}{n} \sum_{i=1}^n \frac{x_i \beta + e_i}{x_i} \\
&= \frac{1}{n} \sum_{i=1}^n \beta + \frac{e_i}{x_i} \\
&= \frac{n\beta}{n} + \frac{1}{n} \sum_{i=1}^n \frac{e_i}{x_i} \\
&= \beta + \frac{1}{n} \sum_{i=1}^n \frac{e_i}{x_i} \\
&\xrightarrow{P} \beta + \mathbb{E} \left[ \frac{e_i}{x_i} \right]
\end{aligned}$$

Recall our assumption that  $\mathbb{E}[e|X] = 0$ . Then:

$$\begin{aligned}
\mathbb{E} \left[ \frac{e_i}{n} \right] &= \mathbb{E}_{x_i} \left[ \frac{1}{n} \mathbb{E}[e_i | x_i] \right] && \text{Using LIE} \\
&= \mathbb{E}_{x_i} \left[ \frac{1}{n} \cdot 0 \right] \\
&= 0
\end{aligned}$$

Then the probability limit of  $\tilde{\beta}$  becomes:

$$\begin{aligned}
\tilde{\beta} &\xrightarrow{P} \beta + \mathbb{E} \left[ \frac{e_i}{x_i} \right] \\
&= \beta + 0 \\
&= \beta
\end{aligned}$$



## Chapter 12

# Summary of Econometrics I

### 12.1 Previous Problem: Hansen II 7.7

Of the variables  $(Y^*, Y, X)$  only the pair  $(Y, X)$  are observed. In this case, we say that  $Y^*$  is a **latent variable**. Suppose

$$\begin{aligned} Y^* &= X'\beta + e \\ \mathbb{E}[Xe] &= 0 \\ Y &= Y^* + u \end{aligned}$$

where  $u$  is a measurement error satisfying

$$\begin{aligned} \mathbb{E}[Xu] &= 0 \\ \mathbb{E}[Y^*u] &= 0 \end{aligned}$$

Let  $\hat{\beta}$  denote the OLS coefficient from the regression of  $Y$  on  $X$ .

- (a) Is  $\beta$  the coefficient from the linear projection of  $Y$  on  $X$ ?
- (b) Is  $\hat{\beta}$  consistent for  $\beta$  as  $n \rightarrow \infty$ ?
- (c) Find the asymptotic distribution of  $\sqrt{n}(\hat{\beta} - \beta)$  as  $n \rightarrow \infty$ .

#### 12.1.1 Part a

We use the OLS formula for  $X$  on  $Y$  and then plug in for  $Y$ :

$$\begin{aligned}
\beta &= \mathbb{E}[xx']^{-1}\mathbb{E}[xy] \\
&= \mathbb{E}[xx']^{-1}\mathbb{E}[x(y^* + u)] \\
&= \mathbb{E}[xx']^{-1}\mathbb{E}[xy^*] + \mathbb{E}[xx']^{-1}\mathbb{E}[xu]
\end{aligned}$$

We know from the assumptions given that  $\mathbb{E}[xu] = 0$ . Plugging this in:

$$\begin{aligned}
\beta &= \mathbb{E}[xx']^{-1}\mathbb{E}[xy^*] + \mathbb{E}[xx']^{-1}\mathbb{E}[xu] \\
&= \mathbb{E}[xx']^{-1}\mathbb{E}[xy^*] + \mathbb{E}[xx']^{-1} \cdot 0 \\
&= \mathbb{E}[xx']^{-1}\mathbb{E}[xy^*]
\end{aligned}$$

This is the linear projection coefficient of  $Y^*$  on  $X$ . So we see that the  $\beta$  coefficient is both the estimator for  $Y^*$  on  $X$  and for  $Y$  on  $X$ .

### 12.1.2 Part b

We apply the OLS estimator to the observed equation:

$$\begin{aligned}
\hat{\beta} &= \left( \frac{1}{n} \sum_{i=1}^n x_i x_i' \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^n x_i y_i \right) \\
&= \left( \frac{1}{n} \sum_{i=1}^n x_i x_i' \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^n x_i (y_i^* + u_i) \right) \\
&= \left( \frac{1}{n} \sum_{i=1}^n x_i x_i' \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^n x_i y_i^* \right) + \left( \frac{1}{n} \sum_{i=1}^n x_i x_i' \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^n x_i u_i \right) \\
&= \left( \frac{1}{n} \sum_{i=1}^n x_i x_i' \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^n x_i (x_i' \beta + e_i) \right) + \left( \frac{1}{n} \sum_{i=1}^n x_i x_i' \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^n x_i u_i \right) \\
&= \left( \frac{1}{n} \sum_{i=1}^n x_i x_i' \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^n x_i x_i' \right) \beta + \left( \frac{1}{n} \sum_{i=1}^n x_i x_i' \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^n x_i e_i \right) \\
&\quad + \left( \frac{1}{n} \sum_{i=1}^n x_i x_i' \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^n x_i u_i \right) \\
&= \beta + \left( \frac{1}{n} \sum_{i=1}^n x_i x_i' \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^n x_i e_i \right) + \left( \frac{1}{n} \sum_{i=1}^n x_i x_i' \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^n x_i u_i \right) \quad (1)
\end{aligned}$$



Now take  $n \rightarrow \infty$ :

$$\hat{\beta} \xrightarrow{P} \beta + \mathbb{E}[x_i x_i']^{-1} \mathbb{E}[x_i e_i] + \mathbb{E}[x_i x_i']^{-1} \mathbb{E}[x_i u_i]$$

Given our assumptions that  $\mathbb{E}[x_i e_i] = 0$  and  $\mathbb{E}[x_i u_i] = 0$ , we find that:

$$\hat{\beta} \xrightarrow{P} \beta$$

So  $\hat{\beta}$  is consistent for  $\beta$  if our covariates are orthogonal to the measurement error term.

### 12.1.3 Part c

Starting from equation (1) in part (b):

$$\begin{aligned} \hat{\beta} &= \beta + \left( \frac{1}{n} \sum_{i=1}^n x_i x_i' \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^n x_i e_i \right) + \left( \frac{1}{n} \sum_{i=1}^n x_i x_i' \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^n x_i u_i \right) \\ \hat{\beta} - \beta &= \left( \frac{1}{n} \sum_{i=1}^n x_i x_i' \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^n x_i e_i \right) + \left( \frac{1}{n} \sum_{i=1}^n x_i x_i' \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^n x_i u_i \right) \\ \sqrt{n} (\hat{\beta} - \beta) &= \left( \frac{1}{n} \sum_{i=1}^n x_i x_i' \right)^{-1} \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n x_i e_i \right) + \left( \frac{1}{n} \sum_{i=1}^n x_i x_i' \right)^{-1} \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n x_i u_i \right) \end{aligned}$$

Assuming that  $\mathbb{E}[eu] = 0$ , we can ignore the covariance term between the two errors when finding the asymptotic variance:

$$\begin{aligned} \sqrt{n} (\hat{\beta} - \beta) &\xrightarrow{d} \mathbb{E}[x_i x_i']^{-1} N(0, \mathbb{E}[x_i e_i e_i' x_i']) + \mathbb{E}[x_i x_i']^{-1} N(0, \mathbb{E}[x_i u_i u_i' x_i']) \\ &= N(0, \mathbb{E}[x_i x_i']^{-1} \mathbb{E}[x_i e_i e_i' x_i'] \mathbb{E}[x_i x_i']^{-1}) + N(\mathbb{E}[x_i x_i']^{-1} \mathbb{E}[x_i u_i u_i' x_i'] \mathbb{E}[x_i x_i']^{-1}) \end{aligned}$$

Let  $Q_{xx} \equiv \mathbb{E}[x_i x_i']$ . Then:

$$\sqrt{n} (\hat{\beta} - \beta) \xrightarrow{d} N(Q_{xx}^{-1} \mathbb{E}[x_i e_i e_i' x_i'] Q_{xx}^{-1} + Q_{xx}^{-1} \mathbb{E}[x_i u_i u_i' x_i'] Q_{xx}^{-1})$$

## 12.2 Maximum Likelihood Estimation: Hansen 10.8

Find the Cramér-Rao lower bound for  $p$  in the Bernoulli model. In Section 10.3, we derived that the MLE for  $p$  is  $\hat{p} = \bar{X}_n$ . Compute  $\text{Var}(\hat{p})$ . Compare  $\text{Var}(\hat{p})$  with the Cramér-Rao lower bound.

### 12.2.1 Finding the Cramér-Rao lower bound

We first note that the Cramér-Rao lower bound is given by  $(n\mathcal{I}_{\theta,1})^{-1}$ . So we know we need to first find the likelihood function. The pmf of a Bernoulli random variable is  $\pi(x) = p^x(1-p)^{1-x}$ . So:

$$\begin{aligned}\mathcal{L}(p|x) &= \prod_{i=1}^n p^{x_i}(1-p)^{1-x_i} \\ \ell_n(p) &= \sum_{i=1}^n \ln(p^{x_i}(1-p)^{1-x_i}) \\ &= \sum_{i=1}^n x_i \ln(p) + (1-x_i) \ln(1-p)\end{aligned}$$

Next, we know we need the score vector, so:

$$\begin{aligned}\frac{\partial \ell_n(p)}{\partial p} &= \frac{\partial}{\partial p} \left[ \sum_{i=1}^n x_i \ln(p) + (1-x_i) \ln(1-p) \right] \\ &= \frac{\sum_{i=1}^n x_i}{p} - \frac{\sum_{i=1}^n (1-x_i)}{1-p}\end{aligned}$$

If we assume that the information matrix equality holds, then we can use the second derivative:

$$\begin{aligned}\frac{\partial^2 \ell_n(p)}{\partial p^2} &= \frac{\partial}{\partial p} \left[ \frac{\sum_{i=1}^n x_i}{p} - \frac{\sum_{i=1}^n (1-x_i)}{1-p} \right] \\ &= -\frac{\sum_{i=1}^n x_i}{p^2} - \frac{\sum_{i=1}^n (1-x_i)}{(1-p)^2}\end{aligned}$$

Now take the negative expectation:

$$\begin{aligned}\mathbb{E} \left[ -\frac{\partial^2 \ell_n(p)}{\partial p^2} \right] &= \mathbb{E} \left[ \frac{\sum_{i=1}^n x_i}{p^2} + \frac{\sum_{i=1}^n (1-x_i)}{(1-p)^2} \right] \\ &= \frac{1}{p^2} \sum_{i=1}^n \mathbb{E}[x_i] + \frac{n}{(1-p)^2} - \frac{\sum_{i=1}^n \mathbb{E}[x_i]}{(1-p)^2} \\ &= \frac{np}{p^2} + \frac{n}{(1-p)^2} - \frac{np}{(1-p)^2} \\ &= \frac{n}{p} + \frac{n(1-p)}{(1-p)^2} \\ &= \frac{n - np + np}{p(1-p)} \\ &= \frac{n}{p(1-p)}\end{aligned}$$

Since the information matrix equality is assumed to hold:

$$\mathcal{I}_{\theta,n} = \frac{n}{p(1-p)}$$

and because our data is *i.i.d.*:

$$n\mathcal{I}_{\theta,1} = \frac{n}{p(1-p)}$$

We have one last step. We must invert this last expression:

$$(n\mathcal{I}_{\theta,1})^{-1} = \frac{p(1-p)}{n}$$

This is the Cramér-Rao lower bound.

### 12.2.2 Variance of the Estimator

We still need to find  $Var(\hat{p})$ . So:

$$\begin{aligned} Var(\hat{p}) &= Var\left(\frac{1}{n}\sum_{i=1}^n x_i\right) \\ &= \frac{1}{n^2}Var\left(\sum_{i=1}^n x_i\right) \\ &= \frac{1}{n^2}\sum_{i=1}^n Var(x_i) \\ &= \frac{1}{n^2}\sum_{i=1}^n p(1-p) \\ &= \frac{n}{n^2}p(1-p) \\ &= \frac{p(1-p)}{n} \end{aligned}$$

Note that this is the Cramér-Rao lower bound.

## 12.3 Method of Moments Estimation: Hansen 11.3

A Bernoulli random variable  $X$  is

$$P(X = 0) = 1 - p$$

$$p(X = 1) = p$$

- (a) Propose a moment estimator  $\hat{p}$  for  $p$ .
- (b) Find the variance of the asymptotic distribution of  $\sqrt{n}(\hat{p} - p)$ .
- (c) Propose an estimator of the asymptotic variance of  $\hat{p}$ .

### 12.3.1 Part a

We begin with the moment conditions:

$$\mathbb{E}[x_i - \mu] = 0$$

$$\mathbb{E}[x_i] = \mu$$

Note that for a Bernoulli random variable  $\mu = p$ . Use the method of moments estimator:

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i$$

$$\hat{p} = \frac{1}{n} \sum_{i=1}^n x_i$$

### 12.3.2 Part b

Using the WLLN, we know that  $\hat{p} \xrightarrow{P} p$ . So we just need to find the variance:

$$\begin{aligned} V &= \mathbb{E}[(x - p)(x - p)] \\ &= \mathbb{E}[(x - p)^2] \\ &= p(1 - p) \end{aligned}$$

### 12.3.3 Part c

This part is really as easy as it seems. Put hats on top of the unknown parameters in our asymptotic variance:

$$\widehat{AVAR} = \hat{p}(1 - \hat{p})$$

## 12.4 Regression Tests: Hansen 13.3 Extended

Take the exponential model with parameter  $\lambda$ . We want a test for  $H_0 : \lambda = 1$  against  $H_A : \lambda \neq 1$ .

- (a) Develop a test based on the sample mean  $\bar{X}_n$ .
- (b) Find the likelihood ratio statistic.
- (c) Find the score test.
- (d) Find the Wald statistic.

### 12.4.1 Part a

We begin with the standard asymptotic t-test:

$$\begin{aligned} T(\mathbf{x}, \mu_0) &= \frac{\sqrt{n}(\bar{x} - \mu_0)}{\hat{s}} \\ &= \frac{\sqrt{n}(\bar{x} - \lambda_0)}{\hat{s}} \\ &= \frac{\sqrt{n}(\bar{x} - 1)}{\hat{s}} \end{aligned}$$

This is our asymptotic t-test.

Note that in this scenario, because we know the true distribution, we can actually calculate the standard deviation:

$$\begin{aligned} \text{Var}(\sqrt{n}\bar{x}) &= n\text{Var}\left(\frac{1}{n}\sum_{i=1}^n x_i\right) \\ &= \frac{1}{n}\sum_{i=1}^n \text{Var}(x_i) \\ &= \frac{1}{n}\sum_{i=1}^n \hat{\lambda}^2 \\ &= \hat{\lambda}^2 \\ &= \bar{x}^2 \end{aligned}$$

Then our t-test becomes:

$$\begin{aligned} T(\mathbf{x}) &= \frac{\sqrt{n}(\bar{x} - 1)}{\hat{s}} \\ &= \frac{\sqrt{n}(\bar{x} - 1)}{\sqrt{\bar{x}^2}} \\ &= \frac{\sqrt{n}(\bar{x} - 1)}{\bar{x}} \end{aligned}$$

### 12.4.2 Part b

Using the formula for the likelihood ratio statistic, and noting that the exponential distribution has a pdf of  $f(x|\lambda) = \frac{1}{\lambda}e^{-\frac{x}{\lambda}}$ :

$$\begin{aligned}
 lr(\mathbf{x}) &= 2 \left[ \ln(L(\hat{\theta}|\mathbf{y})) - \ln(L(\hat{\theta}_0|\mathbf{y})) \right] \\
 &= 2 \left[ \ln \left( \prod_{i=1}^n \frac{1}{\hat{\lambda}} e^{-\frac{x_i}{\hat{\lambda}}} \right) - \ln \left( \prod_{i=1}^n \frac{1}{1} e^{-\frac{x_i}{1}} \right) \right] \\
 &= 2 \left[ \sum_{i=1}^n \left( -\ln(\hat{\lambda}) - \frac{1}{\hat{\lambda}} x_i \right) - \sum_{i=1}^n -x_i \right] \\
 &= 2 \left[ -n \ln(\hat{\lambda}) - \frac{n}{\hat{\lambda}} \bar{x}_n + n \bar{x}_n \right] \\
 &= 2 \left[ \bar{x}_n \left( n - \frac{n}{\hat{\lambda}} \right) - n \ln(\hat{\lambda}) \right] \\
 &\xrightarrow{d} \chi_1^2
 \end{aligned}$$

### 12.4.3 Part c

We first need to find the score evaluated at the null hypothesis of  $\lambda_0 = 1$ :

$$\begin{aligned}
 \left[ \frac{d\ell_n(\lambda)}{d\lambda} \right]_{\lambda=\lambda_0} &= -\frac{n}{\lambda_0} + \frac{n}{\lambda_0^2} \bar{x}_n \\
 &= -\frac{n}{1} + \frac{n}{1} \bar{x}_n \\
 &= n(\bar{x}_n - 1)
 \end{aligned}$$

Assuming the information matrix equality holds, the negative Hessian, and thus the information, evaluated at the null value is:

$$\begin{aligned}
 \mathcal{I}_{\lambda_0,1} &= \frac{1}{\lambda_0^2} \\
 &= 1
 \end{aligned}$$

So the score test is:

$$\begin{aligned}
 T_s(\mathbf{x}) &= \frac{1}{n} n(\bar{x}_n - 1) \cdot 1 \cdot n(\bar{x}_n - 1) \\
 &= n(\bar{x}_n - 1)^2 \\
 &\xrightarrow{d} \chi_1^2
 \end{aligned}$$

### 12.4.4 Part d

Last test to derive. Using the constraint  $\mathbf{g}(\cdot)$  defined above:

$$\begin{aligned}\mathbf{g}(\lambda) &= \lambda - \lambda_0 \\ &= \lambda - 1 \\ &= 0\end{aligned}$$

Now we need the derivative of  $\mathbf{g}(\lambda)$  with respect to the **true**  $\lambda$ :

$$\frac{\partial \mathbf{g}(\lambda)}{\partial \lambda} = 1$$

Using the Wald statistic formula from above:

$$\begin{aligned}T_w(\mathbf{x}) &= n\mathbf{g}(\hat{\lambda}_n) \left( \frac{\partial \mathbf{g}(\lambda)}{\partial \lambda} \mathcal{I}_{\hat{\lambda}_n, 1}^{-1} \frac{\partial \mathbf{g}(\lambda)}{\partial \lambda} \right)^{-1} \mathbf{g}(\hat{\lambda}_n) \\ &= n \left( \hat{\lambda}_n - 1 \right) \cdot \frac{1}{\hat{\lambda}_n^2} \cdot \left( \hat{\lambda}_n - 1 \right) \\ &= \frac{n(\hat{\lambda}_n - 1)^2}{\hat{\lambda}_n^2} \\ &= \frac{n(\bar{x}_n - 1)^2}{\bar{x}_n^2} \\ &\xrightarrow{d} \chi_1^2\end{aligned}$$

This result is the square of our t-test as expected.

## 12.5 Final

These problems should mostly cover the topics that could be on your final. We have gone through regression, consistency, asymptotic normality, MLE, MoM, and standard hypothesis testing (look at confidence intervals before Monday). Know these well and you should be fine on the final.

Some tips: do not leave anything blank. At the very least, sketch out how you would approach the problem. As always I am available for questions should you have any. If you have a question that would be difficult to answer over email, feel free to email me to set-up a time to meet or just check to see if I am in my office.

Good luck on Monday and have a great break!